

# Two symbols, one solution

Saving a handful of photogenic species — or iconic rainforests — is no substitute for a comprehensive plan that deals with climate, economics and the environment together.

**T**he polar bear and the Amazon rainforest, two compelling symbols of the challenges presented by global warming, are again making headlines. For the polar bear, the news is that the United States has finally declared it to be a 'threatened' species as a result of climate-induced loss of sea ice (see page 432). The rainforest, meanwhile, has lost one of its most vocal champions with the resignation last week of Brazil's environment minister, Marina Silva.

Both the polar bear and the Amazon need all the protection they can get. But symbols, by themselves, are just that. What is at stake here is not a charismatic species of bear or one, admittedly vast, forest, but the livelihoods of everyone on Earth and the survival of biodiversity on a global scale.

In the case of the polar bear, US interior secretary Dirk Kempthorne deserves credit for approving the listing in the face of considerable pressure to do otherwise. Quite aside from the Bush administration's scepticism of regulation in general, the case for listing the polar bear was not exactly open-and-shut: international hunting restrictions have led to bear populations that are higher today than they have been in decades. Nonetheless, the scientific evidence for the threat was too strong to ignore.

Kempthorne's decision was delayed for months while the administration drew up regulations to prevent environmental activists using the 'threatened' designation in court to halt energy projects and shut down coal-fired power plants across the country. And the administration was correct to do so. The Endangered Species Act should not be used to sneak broad climate-policy decisions in through the back door. The proper place to make such decisions is openly, in Congress, where a debate on one major climate bill is already scheduled for early June.

In Brazil, meanwhile, where massive deforestation in the Amazon basin is adding its own burden of carbon dioxide to the atmosphere, Silva resigned her ministerial post citing difficulties in implementing federal environmental policy. Indeed, her tenure was marked

by frequent disputes with pro-development forces both in industry and in her own government. The final straw may well have been the Brazilian government's new 'sustainable Amazon plan', which she is widely reported to have opposed. The plan would establish cheap loans to encourage better farming practices; increase aid and other social services for families who rely on logging; and set aside new conservation areas. More controversially, it would also provide for infrastructure such as new roads and hydroelectric dams.

Although Silva's resignation certainly raises questions about the viability of the government's scheme, Brazil's leaders are correct that the Amazon needs some such comprehensive plan. It is condescending and counterproductive to say, as UK newspaper *The Independent* did recently, that the Amazon is too important to be left to the Brazilians. In fact, this region is home to some 25 million Brazilians who need to make a living, and it provides the hydroelectricity that powers much of Brazil's growing economy. Brazil has no choice but to manage it. Indeed, President Luiz Inácio Lula da Silva has promised that his efforts to halt deforestation will continue under the new environment minister, Carlos Minc. A co-founder of the Green Party in Brazil, Minc most recently served as the top environmental regulator for the state of Rio de Janeiro.

The world will be watching to see how this plays out. In the meantime, those concerned about the Amazon — and the polar bear — should keep their focus on the real long-term solution: establishing comprehensive climate-regulatory regimes and providing carbon-free energy sources. If all goes well, tomorrow's industrialists might one day discover that it is profitable to reduce emissions by funding conservation programmes in the Amazon. In doing so, they might even help the polar bear. ■

**"What is at stake here is not a charismatic species of bear or one, admittedly vast, forest, but the livelihoods of billions of people."**

## Trials on trial

The Food and Drug Administration should rethink its rejection of the Declaration of Helsinki.

**L**ater this year, the US Food and Drug Administration (FDA) will adopt new standards for human clinical trials conducted without its advance sign-off in foreign countries. The rules will govern whether data from such trials can be used in applications to market the drug in question in the United States. Although these new standards specify how to run such trials to meet US requirements, they are worryingly silent on key issues relating to human rights, in

contrast with the rules currently in effect. As a result, they could open the way to some ethically fraught decisions.

Take the case of the drug Surfaxin, a synthetic, inhaled version of a lung protein the absence of which is a leading cause of death in premature infants. Back in 2001, the drug's manufacturer, Discovery Labs of Warrington, Pennsylvania, was looking for a suitable location in Latin America to run a trial on the therapy. But rather than compare its product to one of the several effective drugs already available, Discovery Labs was proposing to administer a placebo to the 325 infants in the control group.

The trial was redesigned only after the FDA — and unfavourable media attention — reminded Discovery Labs that a placebo-controlled trial of this type would be deemed unethical in the United States,

and other developed countries, because effective treatments were available. As a result, the control group received alternative active treatments.

The FDA estimates that annually it receives data from around 575 foreign drug trials conducted without its knowledge, more and more of which come from trials run in the developing world. Currently, these trials must comply with the Declaration of Helsinki (or with local country laws, whichever offer the most protection) if sponsors want to use the data to win US marketing approval. The declaration, adopted in 1964, and revised several times since, is today endorsed by medical associations from 85 countries. It is widely considered to be the bedrock of protection for research subjects. Its 1989 revision, which the FDA uses as its present standard, states that any patient in any trial “should be assured of the best proven diagnostic and therapeutic method”.

Yet the FDA announced last month that it will shelve the declaration. Starting in October, the FDA intends to adopt a new standard it calls Good Clinical Practice (GCP), which is modelled on a 1996 document developed by drug regulators and pharmaceutical industry representatives from the United States, the European Union and Japan. Although GCP deals with subject protection, it is in essence a manual on how to conduct rigorous clinical trials, not a human-

rights document. For instance, whereas Helsinki explicitly discourages the use of placebos for serious conditions where proven therapies exist, GCP is silent on this issue. So under the GCP guidelines, the FDA could accept data from Surfaxin placebo trials of the future.

The FDA argues that it should not be bound by Helsinki because the declaration is devised by a group it does not control, and is subject to periodic revisions that could confuse trial sponsors or contradict US law. But it is tempting to conclude that the FDA is dropping Helsinki not because it is changeable, but because the agency disagrees with the way it has been changing — in particular with its constraints on the use of placebos. (The US agency is more favourably disposed to placebo use than, say, its European counterparts.)

It makes sense for the FDA to adopt the GCP standard, giving foreign-based researchers guidelines that should help them generate the best data. But if the FDA jettisons Helsinki, the critical underpinning for such efforts, it risks sending a message that ethical considerations are expendable when research subjects live half a world away. ■

**“The FDA risks sending a message that ethical considerations are expendable when research subjects live half a world away.”**

## The Universe at home

The digitization of astronomy is a transformation and a delight for both amateurs and professionals.

To look through even a small telescope at the greatness of the sky is a heady thing. It is not just the aesthetic delight of stars like grains of sand, or cloud-decked galaxies like tiny hurricanes in seas unseen; nor is it merely the knowledge of immensity that comes with understanding that each grain is a sun bright and ancient, each cloud an unknowable plurality of worlds. It is the sheer cosmic kick that comes from having the rods and cones of your retina stirred by photons that have been travelling for so long that mountains once young have crumbled to the sea. A ray of light that begins in another galaxy and ends in your nervous system is a miracle never to be tired of.

This continuing appeal of amateur astronomy should, light pollution permitting, see children and pensioners in their back gardens and up their local hills for as long as there are telescope makers to satisfy them. Online services such as Google Sky and Microsoft's new WorldWide Telescope allow users to scan the sky at higher resolution and in more wavelengths than amateurs could ever do, yet there is no reason to fear that they will bring those skywatchers indoors for good. Quite the reverse: their on-screen wonders feed the appetite to see for yourself.

Better still, the Internet allows the aggregation of observing time — both for those with telescopes and without. Amateurs following up newly discovered asteroids get the orbital elements from the website of the International Astronomical Union's Minor Planet Center. Galaxy Zoo, a site where a million galaxies await classification, puts the profusion of amateur eyeballs to further use, and has produced not

only an unexpected level of interest but also some sound publications. Harnessing the pattern-recognition skills of people around the world who have no astronomical equipment other than a broadband connection may permit a range of similar projects in the future, as new surveys produce images at ever greater rates.

The Astrometry.net software we report this week (see page 437) offers new ways for the Internet to combine the observations of amateurs and professionals, past and present. It aims to provide the correct spatial and temporal coordinates for any picture of the sky submitted to it, be it a recent CCD file or a glass plate found in Great-Aunt Herschel's attic. Its creators hope, with suitably astronomical ambition, to identify, and possibly assimilate, every image of the sky ever made. In doing so, they imagine discovering new truths about the way the sky changes over time — to recognize the transient, the unexpected, the hitherto unnoticed but nevertheless captured.

This is the sort of totalizing impulse that normally deserves scepticism, if not disquiet. In this case, though, it is hard not to see it as noble. The idea that all the solitary skywatchers are engaged in a single study, linked by ties of knowledge even as they stare upwards on their own, has always had its poetic truth. To make it practically true is a fine aspiration. Walt Whitman's poetic narrator may have

Wander'd off by myself,  
In the mystical moist night-air, and from time to time,  
Look'd up in perfect silence at the stars

— but there is no need to reject the learned astronomers' proofs and figures, charts and diagrams in order to experience that which enriches the soul. Observers of all sorts will soon be able to add to the glories of the endlessly interconnected inner world of the Internet while losing nothing of that precious and primal communion with the cosmos. ■

# RESEARCH HIGHLIGHTS

## Fjord focus

*Nature Geosci.* doi:10.1038/ngeo201 (2008)

*The Hitchhikers Guide to the Galaxy* ascribes the pleasing depth and drama of Earth's fjords to an alien planetary engineer named Slartibartfast. A model produced by Mark Kessler at the University of Colorado in Boulder and his colleagues captures a more plausible sculptural process involving only ice and mountain ranges. The researchers show that a tendency for ice to flow through existing mountain passes deepens these passes, reinforcing the original tendency. This feedback can lead to the carving of kilometre-deep fjords in a million years.

Their model also suggests that once a landscape is equipped with fjords its ice caps will be smaller and more sensitive to climate change, as it is easier for the ice to get away.



G. WILTSIE/NATL GEOGRAPHIC

## MICROSCOPY

### Enter, the nanoscope

*Nature Methods* doi:10.1038/nmeth.1214 (2008)

A high-resolution microscope built in Germany can capture three-dimensional images of proteins within tiny cellular organelles such as mitochondria.

Traditional fluorescence microscopes image a sample 'slice-by-slice' and then assemble those images into a three-dimensional picture. They usually handle slices more than 200 nanometres thick. Alexander Egner and Stefan Hell of the Max Planck Institute for Biophysical Chemistry in Göttingen and their co-workers have created a 'nanoscope' that improves resolution in three dimensions and can image slices measuring about 40 nanometres across.

The researchers used their invention to build up a picture of the distribution of a fluorescently labelled protein called Tom20 in mitochondria. They found that Tom20 forms clusters in the outer mitochondrial membrane.

to three adult female sloths in the tropical forest surrounding the Smithsonian Tropical Research Institute in Panama.

The recordings are the first of their kind from any animal in the wild. The researchers suggest that, because they need to find their own food and keep a look out for predators, other wild creatures may also spend less time in slumber than has been assumed from studies of animals in captivity.

## CHEMISTRY

### Up close and structural

*Science* **320**, 924–928 (2008)

Chemists from the University of Virginia in Charlottesville have developed a rotational spectroscopy technique that allows them to watch as a molecule alters the conformation of its constituent atoms.

Such changes take only picoseconds, although they happened 16 times more slowly than theory had predicted in the molecule that Brooks Pate and his co-workers studied. The team blasted cyclopropane

carboxaldehyde with energy, then followed as it switched between 'syn' and 'anti' isomers, its two stable forms.

Collecting the data took only 52 hours. Using standard methods and apparatus it would have taken about 27 years, the authors estimate.

## MOLECULAR BIOLOGY

### AAAAnswers

*Genes Dev.* **22**, 1141–1146 (2008)

Cells 'tag' newly synthesized RNA with tails of repeating units of adenine in order to make the RNA molecule more stable and prepare it for life in the cytoplasm.

Rebecca Oakey of King's College London and her colleagues report that, for a particular mouse gene, the choice of tagging site correlates with the extent to which the relevant DNA carries methyl groups. This methylation is a form of 'epigenetic imprinting' — a propensity for a particular copy of a gene to be expressed or not that is, itself, inherited.

This is the first evidence that epigenetic imprinting can affect the composition of RNA transcripts in this way.

## POPULATION BIOLOGY

### Keep off the grass

*Biol. Lett.* doi:10.1098/rsbl.2008.0106 (2008)

The dramatic cycling of vole populations with time may be driven by grasses responding to the furry creatures' herbivory. So say Fergus Massey of the University of Sussex in Brighton, UK, and his co-workers, who studied interactions between the vole

## ZOOLOGY

### Less than slothful

*Biol. Lett.* doi:10.1098/rsbl.2008.0203 (2008)

Wild brown-throated three-toed sloths (*Bradypus variegatus*; pictured right) sleep for a mere nine to ten hours a day, much less than the 16 hours of shut-eye observed in captive sloths.

Niels Rattenborg of the Max Planck Institute for Ornithology in Starnberg, Germany, and his colleagues made the discovery by fitting a miniature electroencephalogram (EEG), which measures electrical activity in the brain,



M. FOGDEN & P. FOGDEN/MINDEN PICTURES/FLIPA



*Microtus agrestis* and its main winter food source, *Deschampsia caespitosa*, at four sites in a forest in northern England.

Grasses contained the most silica where the density of voles had been high in the previous spring but was declining during the study period. Meanwhile, little silica was found where the vole population had been low in spring the year before but had since begun to rise.

The team proposes that munching voles prompt grasses to store more silica, which reduces the ease with which voles can digest the plants. So vole growth and reproduction rates fall, and the population tumbles. Grasses then reduce their silica content and the cycle begins again.

## GENETICS

### Self defence

*Science* **320**, 935–938 (2008)

An 'immune system' embedded in *Escherichia coli*'s genome protects the bacterium against injurious genes acquired from other organisms.

A protein called Rho helps *E. coli* produce some RNA molecules because it indicates when to stop making them. Evgeny Nudler of New York University School of Medicine, Max Gottesman of Columbia University Medical Center, New York, and their team exposed *E. coli* to an antibiotic called bicyclomycin that inhibits Rho. Without this stop signal, many genes that viruses and other bacterial species have transferred into the *E. coli* genome were newly expressed, often with toxic effects.

One such gene, acquired from a virus, inhibits cell division. Taken together, the findings suggest that Rho blocks the expression of harmful foreign genes.

## INORGANIC CHEMISTRY

### Towards a noble line

*J. Am. Chem. Soc.* **130**, 6114–6118 (2008)

For decades chemists have known that noble gases can subvert their name by forming chemical compounds. No compound demonstrates this point as emphatically as HXeOXeH, a molecule prepared by Leonid Khriachtchev at the University of Helsinki in Finland and his colleagues.

The compound is almost unique in containing two noble-gas atoms in a single, small molecule, and is possibly the simplest molecule of this type. The structure is like

that of a water molecule with a xenon atom inserted into both of the bonds between oxygen and hydrogen. HXeOXeH forms in a photochemical reaction between xenon and water at 45 kelvin. The researchers hope their finding will be the first step towards designing polymers with alternating xenon and oxygen atoms.

## PLANT SCIENCE

### Monster fruit

*Nature Genet.* doi:10.1038/ng.144 (2008)

Fat, juicy tomatoes may be the norm in modern supermarkets; wild tomatoes can be 1,000 times smaller. Biologists at Cornell University in Ithaca, New York, have identified a major genetic determinant of large tomato size that increases the number of female reproductive organs in a tomato flower, and thus the number of compartments in the fruit.

The determinant, 6–8-kilobases long, is in a gene called *fas*, named by Steven Tanksley and his colleagues. The team crossed tomatoes of varying girth and mapped the genetic region that conferred the tomatoes' compartment number.

The insertion in *fas* is probably a mutation that occurred during tomato domestication; it was not present in 30 lines of the wild tomato from which domestic tomatoes are thought to descend.



MANCAU/PHOTOCUSINE/CORBIS

## HUMAN REPRODUCTION

### Fertile tones

*Evol. Hum. Behav.* doi:10.1016/j.evolhumbehav.2008.02.001 (2008)

American women in the fertile stage of their menstrual cycle speak with voices that are more attractive than those of women who are passing through their infertile stage.

Nathan Pipitone and Gordon Gallup at the University at Albany, part of the State University of New York, recorded the voices of 51 women at four points in their menstrual cycles. Voice samples were then played to 34 male and 32 female participants who rated the attractiveness of the voices.

Participants of both genders scored the voices as more attractive when the women were fertile, suggesting that although women do not obviously advertise their fertility, they unwittingly send subtle cues. Perhaps unsurprisingly, such variation was absent in women taking the contraceptive pill.

## JOURNAL CLUB

Andrea Manica  
University of Cambridge, UK

### A zoologist traces flu across the globe.

In winter, everybody recognizes a stuffy nose, a fever and an achy body as influenza. But experts still grapple with where the flu virus goes during the summer. One theory has it that flu lays low, holding out until the following season in a small number of asymptomatic people. Another idea — that flu strains tend to become extinct locally but shift around geographically — carries more weight. A recent paper by Derek Smith of the University of Cambridge, UK, and his colleagues helped nail the latter hypothesis by plotting the results of antigen-binding assays and genetic sequencing of more than ten thousand viruses on a map (C. A. Russell *et al.* *Science* **320**, 340–346; 2008).

The researchers call this approach 'antigenic cartography'. Their antigenic time charts contain data crunched from the portion of the World Health Organization's enormous 'Global Influenza Surveillance Network' database that details strains classified as 'H3N2' between 2002 and 2007. First, they confirm flu's source-sink dynamics by showing that winter flu strains are more closely related to (and thus more likely to have evolved from) strains found elsewhere than to last season's local contagion. Second, the team pinned down H3N2's spread. Temperate regions are regularly seeded by strains from east and southeast Asia, where many strains circulate continuously and asynchronously in a pattern probably driven by varying climatic conditions.

These findings suggest that close surveillance of emerging strains in east and southeast Asia could enable us to predict those that will later affect the rest of the world. Yet it also poses a question: why do flu strains not return to this region after spending time (and thus evolving) elsewhere? Now that we know where new strains come from, we need to find out why they never go back.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

# Polar bear numbers set to fall

In a long-anticipated decision hailed as a victory by environmental groups, the United States last week declared the polar bear (*Ursus maritimus*) a 'threatened' species. But this heightened protection status may have little bearing on the animals' ultimate fate.

The listing, announced by secretary of the interior Dirk Kempthorne on 14 May, connects the continuing retreat in Arctic sea ice due to global warming with large potential reductions in the polar-bear population. Last autumn, the US Geological Survey concluded that the animals are likely to lose 42% of their summer sea ice habitat by mid-century, cutting the world's polar-bear population — estimated at 25,000 — by two-thirds.

Despite this dramatic projection, researchers note that polar bears range across a variety of nations, each with its own conservation approaches, and a variety of habitats, each of which will be affected differently by climate change. Their fates may vary from place to place, too. "I don't believe the polar bear will go extinct, but in some areas they will be heavily reduced and may disappear," says veterinary biologist Christian Sonne of the National Environmental Research Institute in Roskilde, Denmark. Factors other than global warming compound stress on the bears, including the accumulation in fat of polychlorinated biphenyls and other pollutants that lower reproductive capacity and weaken the immune system.

Projecting the fate of a creature that ranges over more than 25° of latitude is difficult. The polar-bear specialist group of the International Union for Conservation of Nature (IUCN) has identified 19 distinct populations that live in markedly different habitats (see map). "Some populations are clearly in far more trouble than others," says biologist Ian Stirling of the Canadian Wildlife Service in Edmonton, Alberta.

For instance, bears that spend the majority of their time on ice may have to migrate long distances to maintain their lifestyle, an additional stress if food is scarce. But polar bear populations in the Canadian archipelago may be fairly stable in the next few decades, as projections suggest that summer sea ice there will be more persistent.

Still others, such as the southernmost populations around Canada's Hudson Bay, may already be experiencing the effects of climate change. Recent studies have shown that such bears are losing valuable hunting time in the spring, when the animals take in most of the year's energy by

## POLAR BEARS THE WORLD ROUND

Populations of *Ursus maritimus* are hard to pin down, but an International Union for Conservation of Nature group has compiled data on their latest numbers. The years the data were recorded are given in parentheses.



fattening up on nesting ringed seals. West of Hudson Bay, young bears are less likely to survive after earlier sea-ice break-ups, a process which now occurs roughly three weeks earlier than it did 30 years ago<sup>1</sup>. South of the bay, the mass-to-body-length ratio of bears in Ontario has more than halved<sup>2</sup> since the early 1980s.

## Need to adapt

Some bear populations may be able to adapt by spending more time on land, but much depends on how quickly the Arctic ice changes.

"I think it depends on how fast this happens," says biologist Erik Born of the Greenland Institute of Natural Resources in Nuuk. "Polar bears in some sense are behaviourally flexible, but they are also really specialized to hunt on sea ice."

In the face of sea-ice declines, the best way to manage the bear may be to minimize other threats, Stirling says — to protect denning

areas, minimize offshore activities and human traffic, reduce hunting or ensure hunts "move over to bears that are going to die anyway".

That may depend heavily on what circumpolar states do next. The US listing, which was forced by an environmental lawsuit in 2005, places polar bears under the auspices of the powerful Endangered Species Act. But officials wrote the rule in such a way that the 1972 Marine Mammal Protection Act can take precedence. This means that the listing may add no additional limitations to offshore oil and gas drilling. Kempthorne also argued that the new listing could not be used to regulate greenhouse-gas emissions.

No circumpolar state regulates greenhouse-gas emissions specifically to protect the polar bear. Norway, which has had the strongest protections, upgraded the bear's status to 'vulnerable' on its Red List of imperilled species after the IUCN did the same in 2006. But "that doesn't change anything," says Dag Vongraven of the Norwegian Polar Institute in Tromsø.

**"Some populations are clearly in far more trouble than others."**

**HAVE YOUR SAY**

Comment on any of our news stories, online.

[www.nature.com/news](http://www.nature.com/news)

Norway's outright ban on hunting is the only regulatory structure to protect polar bears in that country, he says.

The United States, Greenland (under home rule from Denmark) and Canada permit limited hunting. Russia has outlawed polar bear hunts, but illegal kills are thought to be common, says Vongraven.

Canada is also considering whether to upgrade the polar bear's status. Last month, a government advisory committee announced that it would not recommend raising the bear's status to 'threatened' from 'species of concern', a move that could impact hunting activities. A decision will be made after August, when the group's final recommendations are sent to environment minister John Baird.

### Legal battles

In the United States, the new listing is likely to be challenged. "There will clearly be a series of lawsuits over this that will take a long time to resolve," says Holly Doremus, an environmental lawyer at the University of California, Davis. In particular, she says, exempting federal agencies from consulting with the Fish and Wildlife Service on projects involving greenhouse-gas emissions is unlikely to withstand judicial review. "I think the Bush administration is just trying to kick this to the next administration because they don't know how to deal with it," she says.

In the meantime, prompted by other environmental lawsuits, the Fish and Wildlife Service is considering adding other species — including the emperor penguin — to the endangered or threatened species lists, partly because of threats from climate change.

And polar bears are likely to remain at the top of the international agenda for the foreseeable future. "Certainly the polar bear has become that iconic figure that will hopefully become the rallying point for that kind of discussion to take place," says Lyle Laverty, assistant secretary for Fish, Wildlife and Parks.

Next year, officials in the bear's range states plan to meet in Tromsø, Norway, to discuss management options. It will be the first such meeting in 28 years. ■

Rachel Courtland

1. Regehr, E. V., Lunn, N. J., Amstrup, S. C. & Stirling, I. *J. Wildl. Mgmt* **71**, 2673–2683 (2007).
2. Obbard, M. E. et al. *Temporal Trends in the Body Condition of Southern Hudson Bay Polar Bears* Climate Change Research Information Note Number 3, Ontario Ministry of Natural Resources (2006).

See Editorial, page 427.

## Whales are on the rise

Humpback whale numbers in the northern Pacific Ocean have ballooned to nearly 20,000, the largest population seen since the majestic mammals were hunted nearly to extinction half a century ago.

The number of humpbacks hit an all-time low of 1,400 or even lower by 1966, when their hunting was banned internationally. The new census, from one of the largest whale studies ever undertaken, shows that the animals have rebounded much better than expected.

"We had no idea the population could have grown this high," says John Calambokidis, a biologist at Cascadia Research Collective, a non-profit environmental research institute in Olympia, Washington, and a principal investigator on the study.

But cetologists are concerned about the estimated 900 humpbacks that migrate to the western Pacific. This subpopulation may be being hunted illegally, with some getting entangled in the nets of fishermen. Still, researchers say that the western Pacific population is increasing at more than 6% per year — roughly the same rate as humpbacks in other regions.

The three-year study, called SPLASH (Structure of Populations, Levels of Abundance and Status of Humpbacks), involved more than 400 researchers from 10 nations. Its US\$3.7-million price tag was paid for with funding from the US National Oceanic and Atmospheric Administration (NOAA), the Canadian government and private sources. It used everything from ocean-going research ships to motorized outrigger canoes to identify whales by their fluke markings, then monitor them from their feeding grounds off Canada and the Aleutian Islands to their winter and breeding grounds off Hawaii, Latin America and Asia.

"This is a great candidate to show the success of conservation programmes," says

Jay Barlow, a marine mammalogist with the NOAA Southwest Fisheries Science Center in La Jolla, California, and a study leader. The project was conceived in 2002 when US laws such as the Endangered Species Act and the Marine Mammal Protection Act were under attack by Republicans in Congress.

The report's findings may open a new dialogue about the study and regulation of humpbacks under the auspices of the International Whaling Commission. And there will probably be talks about re-evaluating the humpback's current classification as endangered.

Barlow says that revising the protection status to 'threatened' may be reasonable for the eastern-Pacific population, but that western-Pacific whales should remain listed as 'endangered'. "This study will open a discussion, which will be a long one," he says.

Japan continues 'scientific whaling' attempts on a separate population of humpbacks in the southern Pacific. Last year, the country had planned to kill nearly 1,000 humpback, fin and minke whales in the area, but international pressure reduced the take to about 550 minke.

SPLASH also intends to furnish details about the humpback population structure, including the animals' loyalty to certain feeding or breeding regions and how this affects their survival.

Almost 8,000 humpbacks were individually catalogued, with tissue samples taken from more than 6,000 of these for DNA analysis. Already, SPLASH has revealed the existence of an unknown wintering and breeding ground — a refuge that researchers haven't yet located, but is probably in the middle of the Pacific Ocean. DNA records may play an important part in locating the area. "Finding that will be a fun project," says Barlow. ■

Rex Dalton



Long tale: Humpback whales seem to be on the route to recovery.

J. BARLOW



# Meeting urges scientists into politics

Scientists who think the world would be a better place if more of their sort held public office are being urged to help make the changes they want and offered advice on how to do so.

On 10 May, the non-partisan organization Scientists and Engineers for America (SEA) held a workshop in Washington DC to tell scientists what it takes to run for public office — and how to go about it. Around 75 scientists, science teachers, science-policy experts and other interested parties gathered on the campus of Georgetown University to explore the transition from scientist to politician. Some were aspiring politicians. Others hoped to contribute to the inner workings of political campaigns. All of them learned how difficult it is to translate a scientific career into a leadership role in politics.

Early insight came from Kevan Chapman, the communications director for Vernon Ehlers — a physicist and Republican who represents a portion of Michigan. As a political candidate, Ehlers has always touted his scientific background. During his 15 years in Congress, he has emphasized that the greatest asset of his scientific training is his reasoned, rational way of approaching complex problems.

Playing up a training in rational thinking could be a greater benefit in the political arena now than in the past, says Joe Trippi, architect of the innovative online fund-raising effort that fuelled Howard Dean's 2004 campaign for the Democratic presidential nomination. American citizens, Trippi says, are ready for an end to "transactional politics", of the 'Vote for me, and I'll cut your taxes' approach. "No deal in the world," explains Trippi, will bring about an end to global warming, so scientists-turned-politicians should underscore that they have been trained to find solutions, not to trade vacuous political barbs with the opposition.

Sarah Mullins, a chemist at 3M in St Paul, Minnesota, who helps the American Chemical Society to lobby Congress, found Trippi's remarks energizing. "He actually put out a vision of scientists in electoral politics," she said. "It just wasn't the same old 'we need people to be good leaders.'" Mullins has no plans to run for office right now, but she intends to use Trippi's comments to try to convince fellow chemists that they can have an important role in politics.

Still, would-be candidates face numerous practical considerations. The time commitment involved and the importance

**"It is best to clear your plate and focus on running for office."**



For chemists to get to the top, like this one did, is rare.

of establishing oneself in the community can be challenging for scientists whose schedules are often inflexible and who move frequently, especially early in their career.

"It is best to clear your plate and focus on running for office," says Louis Lanzerotti, a physicist and member of the National Science Board who also served as mayor of Harding Township in New Jersey. Lanzerotti says he is often asked for scientific advice out of his field; recently, as a consultant on one project, he calculated the costs and benefits of artificial athletics fields, and the possibility of turf-associated chemicals leaching into nearby streams.

Speaker David Westerling, a civil engineer, agrees that local issues are important, saying that he had his first success by looking at community projects such as drainage, wetlands

and athletics facilities. Westerling told the participants to think about how they could take advantage of their academic contacts, although with caution. Westerling himself experienced a backlash when trying to raise money for his first campaign via his alumni association.

Indeed, some might find that scientist colleagues frown on their political aspirations. Physicist Michael Lubell, director of the American Physical Society in Washington DC, says that his Yale colleagues were puzzled when he started canvassing for votes for the successful Lyndon B. Johnson presidential campaign as a graduate student in 1964. "But," he said, smiling, they "eventually recognized that I had political contacts they could take advantage of".

In July, the SEA plans to start an online advice forum for scientists interested in running for office.

**Gene Russo**



# Sterile mosquitoes near take-off

Malaysia is looking to battle dengue fever by releasing mosquitoes that have been genetically engineered to be sterile. Although these efforts have stirred public concern, the country's Academy of Sciences is likely to recommend the strategy to the government within a month.

In April, the Institute for Medical Research in Kuala Lumpur indicated that it might release millions of male *Aedes aegypti* mosquitoes that have been genetically modified to produce offspring that die in the larval stage. The release of enough of the sterile males would theoretically swamp fertile wild-type competitors and crash the population.

This 'sterile insect technique' has been successful in the past, for example in eliminating the medfly from California and the parasitic screw worm from the United States and much of central America. But those insects were sterilized using radiation, which doesn't work as well on mosquitoes. Irradiated mosquitoes are unable to compete with wild-type males to mate with females.

Scientists in Malaysia have been working with mosquitoes created by Oxitec, a company based in Oxford, UK, and founded by University of Oxford geneticist Luke Alphey. Oxitec integrated a genetic element, LA 513, into the DNA of the mosquitoes. This genetic modification kills any offspring in the larval stage if they are not fed the drug tetracycline. In the lab, the mosquitoes are fed tetracycline and grow in the millions. In the wild, the modified gene kicks in, and, in theory, would be able to crash the local *A. aegypti* population (H. K. Phuc *et al.* *BMC Biol.* doi:10.1186/1741-7007-5-11; 2007).

Between September and December of last year, the Institute for Medical Research, part of the Malaysian health ministry, evaluated Oxitec's RIDL-513A strain of *A. aegypti* in what Alphey describes as "the most realistic environment in which engineered mosquitoes have ever been tested". The engineered mosquitoes did well in competing with wild-type males, mating at the same rate with females. "That would certainly be a step forward over radiation," says Austin Burt of Imperial College London, who works on genetically modifying mosquitoes for malaria control.

Local environmentalists reacted with

alarm to media reports last month that the strain would be released on Pulau Ketam, an island fishing village a few dozen kilometres from Kuala Lumpur. The next day, the government issued a press release saying, "Such a release will never be carried out without the proper clearance of the relevant authorities." Researchers from the Institute for Medical Research did not respond to requests for comment.

Gurmit Singh, an environmentalist and chairman of the non-profit Centre for Environment, Technology and Development in Petaling Jaya, says the main problem is that the government never made public details of the long-term potential for ecological disturbance. "How are the mosquitoes produced, and what's the possibility that the mutation could spread?" Singh asks.

Burt notes that people shouldn't be worried because the mosquitoes are designed to die out rather than spread. In his unrelated work, Burt is trying to modify genes to make it difficult for mosquitoes to pass malaria to humans.

Unlike Oxitec's mosquitoes, Burt's would need to have a selective advantage for the gene to spread. Burt says that changing the gene pool is "not something you do lightly", but that he hopes to have a mosquito ready for trials in the next five years. Burt's project has, like Alphey's, received money from the Bill & Melinda Gates Foundation, which lists as one of its Grand Challenges for Global Health: "Develop a genetic strategy to deplete or incapacitate a disease-transmitting insect population."

Malaysia's field trials could go ahead long before Burt's. Alphey met with the Malaysian Academy of Sciences on 16 May and says the meeting went well; C. P. Ramachandran, who chairs the relevant committee and has decades of experience in tropical disease research, said "the science is excellent". He adds that, "any risks related to genetically modified organisms must be balanced against the potential benefits," noting that Malaysia has tens of thousands of dengue cases each year.

According to the World Health Organization, spread of dengue since the 1970s has put 2.5 billion people at risk, with an estimated 50 million cases each year. ■  
David Cyranoski

**"Any risks related to genetically modified organisms must be balanced against the potential benefits."**

## ON THE RECORD

**"The word God is for me nothing more than the expression and product of human weaknesses."**

Albert Einstein, in a 1954 letter that sold last week at auction in London for £170,000. Richard Dawkins was one of the losing bidders.

## SCORECARD



**Engineering for artists**  
Engineers say they have worked out

how to safely reconstruct Henry Moore's giant 1980 sculpture *Arch*, which was dismantled in 1996 after cracks started to appear. It needs better bolting and firmer foundations, they say — although they haven't worked out the details yet.



THE HENRY MOORE FOUNDATION



**Art for engineers**

The Habog nuclear-waste storage facility has been prettied up by artist William Verstraeten, who painted it orange, with  $E=mc^2$  and  $E=h\nu$  written along the side. The artist says he chose orange because it's halfway between red-for-danger and green-for-safety.



M. VAN DEN BERGH/HOLLANDSE HOOGTE

## ZOO NEWS

**Returned from the choir invisible**

The fossil of a 55-million-year-old parrot has been found in Denmark. Officially named *Mopsitta tanta*, it will forever be known as a 'Norwegian Blue' (despite there being no evidence of its colour) after Monty Python's parrot sketch. This one, too, is clearly dead, but not "pining for the fjords" — Norway didn't get its fjords until more than 50 million years later.

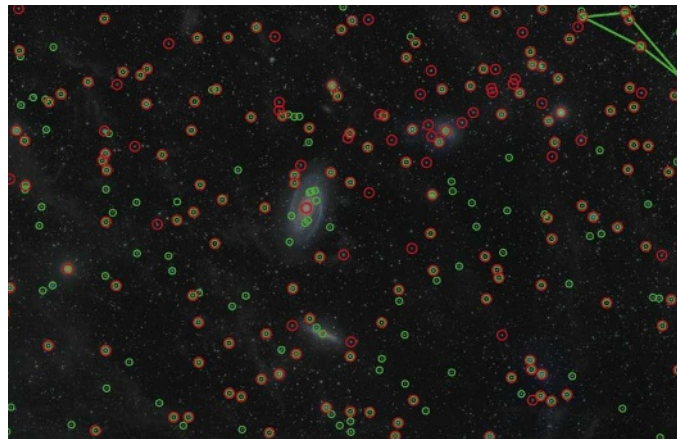
Sources: UPI, Imperial College London, World Nuclear News, AlphaGalileo



**SOCIAL NETWORKING GETS MEDIEVAL**  
Researchers give a French province the 'Facebook' treatment.  
[www.nature.com/news](http://www.nature.com/news)

PUNCHSTOCK

J. GALLEGO



ASTROMETRY.NET

Left, the sky around galaxy M81; right, Astrometry.net uses a recognized shape (top right) to line up stars in the image (red) with those in its memory (green).

# No star left behind

An open-source software project could help unify every existing astronomical image into a single data set.

Everyone knows the experience: a pile of old photos in a dusty attic — or more likely now on a dusty disk drive — with no indication of who is in them or when they were taken. As long as the subjects are astronomical features, though, there's now an answer at hand.

Astrometry.net is an open-source software project, run out of the University of Toronto in Canada and New York University, which aims to recognize any starscape and place it in its proper coordinates within seconds — specifying not just which patch of the sky is shown, but when. By using the small relative motions of stars over time, the project's designers hope to date any picture to within a year.

The project should regularize astronomical data, making it possible to combine images from both professionals and amateurs into easily used databases. "We'd love to touch every astronomical image ever taken," says David Hogg, a project leader and an astrophysicist at New York University. The system is currently undergoing initial alpha-testing by astronomers at various observatories, but there are plans to start accepting images from the public by September.

Although astronomers try to use a universal coordinate system to chart the sky, the reality is a messier business: telescope pointing mechanisms are idiosyncratic and not always properly used. This can make data from different observatories difficult to compare. Astrometry.net solves the matching problem by choosing bright, four-star constellations from the pixels in a submitted image. It then looks for the same shapes in an index of 800,000 such constellations constructed from all-sky surveys catalogued by the US Naval Observatory.

In a test, the software gave the correct

coordinates for all but 451 of 336,554 images from the Sloan Digital Sky Survey, a systematic map of a quarter of the sky. The team has focused on speed and reliability: all of the matches were made in less than a second, and there were no false positive matches.

Tim Axelrod, the data management project scientist for the Large Synoptic Survey Telescope, says it's the most robust calibration system he has seen, and he plans to use it when his telescope begins its surveys in 2014 (the system is powerful enough to survey the entire night sky in three days).

As the Naval Observatory catalogues were made from multiple sky surveys done about 30 years apart, they document tiny shifts in the positions of nearby stars, called their proper motions. This means that, once Astrometry.net has its match, it can turn back the clock, rewinding the proper motions to the best match possible for the pattern documented in a submitted image. Astronomers will thus be able to pick out features that have changed over time, such as Kuiper-belt objects in the Solar System and supernova precursors.

Jonathan Grindlay, a researcher at Harvard University who is interested in transient phenomena, is eager to use the system. He runs a programme aimed at digitally scanning more than half a million photographic plates dating back to the 1880s that, between them, cover every inch of the sky between 500 and 2,000 times. Astrometry.net should go a long way to sorting out exactly what each plate records, and when it was taken. Once it does, a century of the skies will be open to

scrutiny for changes on timescales of decades.

Eventually, Hogg would like the software to recognize any unusual feature as part of the system's routine service. But there are pressing wrinkles to iron out first. An image needs enough four-star constellations for the software to make a match, and so the system has trouble with pictures that cover less than one ten-millionth of the sky. This is an issue for images from instruments such as the Hubble Space Telescope, which looks very deeply into tiny patches of the sky. And to get the most out of the system, the software will need to match images, not just in space and time, but also in brightness

— the apparent brightening and dimming of a star, for instance, would have to be referenced to something.

But as the software improves, it could become the standard for the entire

community, which is trying to supply streams of data to portals such as the planned international 'virtual observatory'. Already, the US National Optical Astronomy Observatory in Tucson, Arizona, plans to be using Astrometry.net by the end of the year to calibrate the 12 terabytes of data it collects annually from its 12 telescopes.

If it ends the laborious days of calibrating astronomical data by hand, Astrometry.net would be saving not only lost photons, but also precious time and money, says Hogg. "The amount of money this would save the astronomical community is immense," he says. ■

**Eric Hand**

See Editorial, page 428.

For further information, see <http://tinyurl.com/6y3bkv>



## Q&amp;A

# Poland tackles science like a business

Polish leaders were disconcerted in January, when the nation's scientists came away empty-handed from the first round of applications for the European Research Council's starting-grant competition. The country is also performing poorly in other European Union research programmes (see page 558). Science minister **Barbara Kudrycka** explains how the Polish government plans to reform the country's science and higher-education system.

## What are the strengths of the existing Polish science system?

We are among the top 20 science nations in fields such as physics, mathematics, chemistry and space. However, we are missing the opportunity to translate this potential into concrete business solutions. That is what we call the innovation gap, and it can best be seen in the number of patents granted in these areas.

But even in applied research there are fields where Polish researchers' performance is among the best in the European Union (EU). For instance, our share of grants awarded through the European research programmes in thematic priorities such as security is outstanding.

## What changes have been proposed?

The biggest change is introducing competitive mechanisms into the system. We want to implement them in the next few years wherever possible — from the distribution of research grants to the development of academic careers. One concrete idea is to create 'flagship' universities.

We also propose abandoning the *Habilitation* [a post-PhD qualification usually required for progress in academia].

## How will science in Poland benefit from the changes?

Well, I would like to say that these changes will lead to Polish Nobel prizewinners, but that would be an overstatement. Our aim is to create excellent conditions for students and scientists in order to create a solid foundation for a knowledge-based society. What we can do from the political point of view is to create an appropriate framework and provide suitable incentives to make science flourish. Whether it does so is a matter for the scientific community itself.

The level of financing dedicated to research and higher education is meant to be increased substantially in the next few years. But these additional funds have to be spent effectively. That is why the whole system has to be better organized, with the ministry playing a strategic part and two funding agencies dealing with cutting-edge applied and basic research.



## How much extra money will become available, and when?

Currently, budgetary expenditure on research and development (R&D) and higher education amounts to little more than 1% of gross domestic product (GDP). We would like to increase funding by 0.15% of GDP per year until 2013. What is equally important is attracting private funds. Today Poland has one of the lowest R&D expenditures financed by the business sector in the whole European Union. We are working on new legislative measures that will make investment in research more attractive for private companies.

## How will the money be distributed?

Our National Science Council will provide funds for frontier research on a competitive basis. Those parts of the system that become competitive will receive extra funding. We want to reduce expenses where the results — scientific, economic or academic — are not evident whatsoever.

## Are you concerned that academia will try to block the reform?

Everyone agrees that we should not only prevent talented young scientists from leaving our country but also attract them from abroad. At present, if an experienced professor from Oxford or Sydney wants

to take up a post at a Polish university, they have to undergo a long and complex evaluation procedure simply because they don't have the proper scientific degree. Parts of the academic community are unenthusiastic about the proposed changes, such as abandoning the *Habilitation*. But *Habilitation* does not exist in most countries. If we want to become attractive we have to be open.

## Are there any particular fields of science that you think should become priorities?

The first priority will be advanced energy technologies. Work on strategic programmes for information and communication technologies and health are also to be concluded very soon.

## How will the reform affect the role of the Polish Academy of Science?

Academy institutes should focus their research on selected areas of science that correspond to our national strategic and priority areas. Some of these institutes are outstanding, some are not that good. We plan to audit the institutes, after which only the best will survive.

## What can be done to make Poland more attractive to young scientists, from Poland and from abroad?

This is a crucial problem for us. Abandoning the *Habilitation* could be helpful, but it is not enough. We would like to impose a system in which each position will be filled in an open, competitive, international process. Attractive postdoctoral fellowships should also be more accessible.

## Poland's participation in EU-funded research is modest at best. How can it be improved?

Polish scientists often don't meet the direction and criteria of EU-funded research. It is not because of the quality, but the subject of the proposed projects that they are rejected. To gain more European grants, we need to modify the direction of our research so that it will respond to EU priorities. ■

Interview by Quirin Schiermeier



**TRIAL ON HOLD**

Drug agency freezes first potential test of embryonic stem-cell-derived therapy.  
[www.nature.com/news](http://www.nature.com/news)

D. SCHARF/SPL

# US plans more primate research

Scientists in the United States are planning for an increase of non-human primate research.

Currently, the National Institutes of Health (NIH) funds eight National Primate Research Centers with a total of about 26,000 animals. But several factors are expected to drive demand, among them the failure last year of an HIV vaccine candidate being trialled by the pharmaceutical company Merck. Such failures have underscored the need for more non-human primate research to answer basic questions about the virus and to develop new vaccine concepts.

"We fully anticipate that the animal model will have a resurgence of interest and importance because we need it to answer some of those fundamental questions," says Anthony Fauci, director of the NIH's National Institute of Allergy and Infectious

Diseases in Bethesda, Maryland.

A greater focus on clinical research is also expected to boost primate work. For instance, on 18 May, researchers from the Yerkes National Primate Center in Atlanta, Georgia, announced that they had made significant progress towards a transgenic model of Huntington's disease in rhesus macaques engineered to carry the genetic defect responsible for the disease (see S.-H. Yang *et al. Nature* doi: 10.1038/nature06975; 2008).

Five of the eight national primate centres are also located near institutions that have received grants specifically to bolster clinical and translational science. "We want to go from basic research forward into preclinical and clinical models, and animal models are a very important part of that type of development," says Barbara

Alving, director of the National Center for Research Resources in Bethesda.

But meeting the demand is a complicated issue. India, the preferred country of origin for the animals, has a long-standing ban on exporting rhesus macaques. Breeding more will take years, and it is not yet clear how many additional animals will be needed, because scientists have not yet told the NIH exactly what research must be done. With agency budgets staying flat in recent years, the primate centres have already formed a consortium to pool resources and make sure they breed enough animals to meet the needs of NIH-funded scientists.

Fauci says the NIH will be consulting with HIV researchers and scientists over the coming year to solidify its plans. This month, his institute solicited comment on the 'highly innovative strategies to prevent HIV transmission', or 'HIT-IT' initiative, a programme that is likely to lead to more non-human primate work. ■

**Erika Check Hayden**



**Rhesus macaques are often used for primate research.**

R. T. NOWITZ/PHOTOTAKE/NEWS.COM

## Parkinson's researchers join forces with gene tester

The Parkinson's Institute and Clinical Center in Sunnyvale, California, last week announced an alliance with 23andMe, a gene-testing company in Mountain View, California. The initiative aims to improve methods for collecting large amounts of consistent data on DNA, lifestyle, family history and environmental exposures, to unpick the factors that contribute to Parkinson's disease.

The current initiative, funded by the Michael J. Fox Foundation for Parkinson's Disease, will gather information from 150 people — half with and half without the disease. Researchers will use questionnaire data from the group to develop robust surveys that can be hosted online by 23andMe. The idea is that future 23andMe clients will be offered the chance to take these surveys, so they can easily add their genetic and lifestyle data to the project.

"It could be a real game change" for disease epidemiology, says Bill Langston, director of the Parkinson's Institute. But he acknowledges that it will only have a big impact if many people choose to pay for the genetic tests and participate in studies.

## Europe considers plans for manned spacecraft

The European Space Agency (ESA) is mulling over the idea of turning its latest cargo ship into a manned spacecraft.

Launched for the first time in March, the Automated Transfer Vehicle is a 20-tonne vessel for carrying food, water, oxygen and experiments to the International Space Station. It is designed to burn up on re-entry after its mission is completed. But later this week, the ship's manufacturer, Paris-based aerospace giant EADS Astrium, is expected to announce detailed plans for turning it into a manned spacecraft. The design would include a re-entry capsule for three people



An artist's impression of the Automated Transfer Vehicle fitted with a return capsule.

## Cancer forces Tasmanian devil onto endangered list

The Tasmanian devil (pictured), long beleaguered by a deadly and contagious form of face cancer, is slated to be listed as an endangered species this week.

A spokeswoman for Tasmania's primary industries minister says that the animal will be listed as an endangered species by state officials on 21 May.

The carnivorous marsupial, which was once hunted for food and to protect livestock, was first protected by law in 1941, and listed as vulnerable in 2006. Cancer is thought to have cut the population by up to one-half between 1995 and 2005.

and new safety features.

ESA is taking the proposal seriously, says Manuel Valls, head of policy and planning for the agency's human spaceflight directorate in Noordwijk, the Netherlands. At present, Europe depends on the United States and Russia to transport people into space. Valls says the new craft would probably cost several billion euros to develop. ESA's governing council is expected to review the proposal in November, along with other options.

## British parliament backs hybrid embryos

UK politicians have rejected calls to ban the creation of animal-human hybrid embryos. In a key vote on 19 May, parliamentarians voted by 336–176 in favour of allowing the creation of 'cybrid' embryos, those created from human DNA and an empty animal egg. They also favoured the creation of 'true' genetic animal-human hybrid embryos, purely for research purposes, by the narrower margin of 286 to 223.

Parliamentarians were given a 'free' vote, rather than being whipped along party lines, on the most controversial aspects of a raft of proposed new fertility legislation. The entire bill, after further votes on aspects such as Britain's current abortion limit of 24 weeks, will be voted on as a whole in coming months.

## Cosmic dust hides true brightness of Universe

The Universe may be brighter than thought: perhaps twice as bright as seen from Earth.

Interstellar cosmic dust absorbs some starlight, but quantifying this has proved difficult. Researchers from the United Kingdom, Germany and Australia looked

closely at how dust obscures light in some nearby galaxies, depending on various factors including the size, shape and inclination of the galaxy. They used these data to make a better model of how dust absorbs light, and applied the model to a catalogue of 10,000 galaxies.

The result suggests that dust is blocking nearly half of this light from our view, making the true brightness of the nearby Universe nearly twice that seen from Earth (S. P. Driver *et al. Astrophys. J. Lett.* 678, L101–L104; 2008). The researchers say that their findings could force a revision in mass estimates for dust-shrouded galaxies, and a reconsideration of models of galaxy formation in the early Universe.

## NOAA chief backs bid for climate-change agency

The chief of the US National Oceanic and Atmospheric Administration (NOAA) has called for the creation of a National Climate Service to manage and disseminate information about global warming.

Like the National Weather Service, the service would fall within NOAA and be a repository for federal research that would be accessible to agencies and the public alike. NOAA administrator Conrad Lautenbacher says that having a central scientific organization to collate this information could help to depoliticize climate research. The service would not have any regulatory authority.

The Senate Committee for Commerce, Science and Transportation passed legislation by John Kerry (Democrat, Massachusetts) to create such a Climate Service last December. Lautenbacher says that the White House supports the idea in principle. NOAA plans to seek funding to organize the service in the 2010 budget proposal.





20TH CENTURY FOX/PHOTOFEST

## Line-ups on trial

A major, but flawed, study of identity parades, or line-ups, has set science and the police at odds. **Laura Spinney** investigates.

In 1981, 22-year-old Jerry Miller was arrested and charged with the kidnap, rape and robbery of a woman in downtown Chicago. After two eyewitnesses to the crime picked him out of a line-up, and the victim identified him at his trial, he was convicted and sentenced to 45 years in prison. In March 2007, however, semen on the victim's clothes was subjected to DNA testing and found not to belong to him. His conviction was quashed a month later. He had spent more than 24 years in jail.

Miller's story is not unique: his was the 200th conviction to be overturned in the United States on the basis of DNA evidence. According to the Innocence Project, an advocacy group based in New York that campaigns to overturn wrongful convictions, and that took on Miller's case in 2005, the number today stands at 216, of whom 16 spent time on death row. Mistaken eyewitness testimony was a factor in more than three-quarters of those cases.

Those exonerations have highlighted the fact that police procedures for eyewitness

identification in the United States are out of step with almost three decades of psychological research, and have triggered a row over whether the procedures should be reformed. As the situation is similar, or worse, in most other developed countries, the rest of the world is watching closely.

The traditional US procedure is familiar to any fan of television cop shows. Witnesses are presented with a line-up that includes both the suspect and a number of innocent people, or 'foils', and are asked to identify the perpetrator. In the early 1990s, however, when the confidence of the justice system had been badly shaken by the first wave of DNA exonerations, the then attorney-general, Janet Reno, invited experts to form a working group to address how this method could be improved.

The group immediately homed in on the fact that most line-ups are overseen by the case's investigating officer, who knows the suspect's identity. For scientists, this is a major error: even something as seemingly objective as a clinical trial can be affected if the nurse who

administers the injection knows whether the syringe contains a drug or a placebo. It is all but impossible for an experimenter — or an investigating officer — to avoid giving away the 'right' answer through body language, tone of voice or other such unconscious hints. "I have argued for years that the more important reform is for line-ups to be conducted double blind," says Gary Wells, a psychologist at Iowa State University in Ames and a member of Reno's original working party. Witnesses should also be told that the perpetrator may not be in the line-up so that they do not feel obliged to identify someone. In every one of the DNA exonerations that involved mistaken identity, says Wells, the witness had picked the suspect: "It's just that the suspect was innocent." Although the real perpetrator was not in the line-up, the witness somehow ended up picking the person the detective had in mind.

The working group's most important recommendation was that line-ups should be conducted in a double-blind fashion, so that neither the witness nor the official overseeing



the procedure would know who the suspect was. The group also recommended that the suspect and foils be presented sequentially rather than simultaneously, and that the witness be asked to make a decision after each one rather than waiting until the end. This would encourage witnesses to compare each individual to their memory of the offender, rather than to one another — a method the scientific data suggested was more likely to produce a correct identification.

In light of these recommendations, several US states reformed their eyewitness identification procedures. However, according to James Doyle of the John Jay College of Criminal Justice in New York, the recommendations encountered pockets of resistance, one of which was in Illinois. Illinois also happens to have one of the highest DNA exoneration counts in the United States, 28, which in 2000 led the state to impose a moratorium on the death penalty until the causes of its wrongful convictions had been identified and eliminated. That moratorium is still in place. Nonetheless, Illinois police tended to view the working group's guidelines not as safeguards against wrongful convictions, but as hopelessly academic and impractical. So in 2003, the Illinois State Police commissioned its own study<sup>1</sup> to test line-ups under real-world, field conditions, and put it under the direction of lawyer Sheri Mecklenburg, general counsel to the Superintendent of the Chicago Police Department.

### Crime watch

The inherent problem with field studies in this area — and the reason that so few have been done — is that there is no way of knowing for sure whether the suspect is guilty. In a mock crime staged before witnesses in a lab, experimenters know that the suspect is the perpetrator. But in the real world, all that can be said for certain is that the foils did not do it. Hence, the measure that experimenters pay most attention to is how often witnesses identified a foil as the perpetrator.

With this in mind, and with the cooperation of two psychologists and three of the state's police departments — Chicago, Evanston and Joliet — Mecklenburg's Illinois Pilot Program spent a year conducting some 700 eyewitness identifications. Some of the procedures were non-blind and simultaneous; the rest were double-blind and sequential. Both conditions were a mix of live line-ups and photo arrays. The team found that the double-blind, sequential technique produced higher rates of

foil picks — that is, clear errors — and lower rates of suspect picks than the traditional, non-blind line-up. Mecklenburg concluded that the existing system should not be changed.

That conclusion, announced in 2006, immediately sowed confusion in the dozen US states that had been considering updating their procedures. One state, New Jersey, and several smaller jurisdictions had already passed reforms, which remain in place. But some states put their reforms on hold, or even threw them out. Horrified, psychologists pointed to deep methodological flaws in the study, which was never submitted to peer review. According to the psychologists, the decision to change two variables at the same time — blind/non-blind and simultaneous/sequential — made it impossible for Mecklenburg and her colleagues to draw any conclusions from their data. "If one of my undergraduates came to me with that experimental design, I would say 'Go away and do it correctly,'" says Daniel Wright, psychologist at the University of Sussex in Brighton, UK.

Writing in the journal *Law and Human Behaviour*<sup>2</sup>, psychologists Daniel Schachter of Harvard University in Cambridge, Massachusetts, and Nobel laureate Daniel Kahn-

man of Princeton University in New Jersey and their colleagues were unanimous that the design flaw "has devastating consequences for assessing the real-world implications of this particular study". After that critique was published, a

handful of US jurisdictions did reform their procedures, and others set up task forces to consider the options. But for Stephen Saloom, policy director at the Innocence Project, that wasn't enough. "Given that the vast majority of jurisdictions nationwide have yet to adopt the reforms, and that eyewitness identification has a role in a set of crimes far greater than those in which DNA can prove innocence or guilt, it is of great concern that the majority of jurisdictions are simply sticking with what they have always done," he says.

Last year, the National Association of Criminal Defense Lawyers in Washington DC and the MacArthur Justice Center at Northwestern

University in Chicago, Illinois, both of which back reform, sued the police departments that took part in the pilot programme to release their data so that they could have them reanalysed. So far the departments have refused, and Mecklenburg calls the lawsuit "a waste of energy". She says it has only alienated the police, making it less likely that they will cooperate with field studies in future.

Regarding the study itself, Mecklenburg has stood her ground. She argues that pristine lab studies capture none of the tension and gravity of an actual line-up. Moreover, although admitting that the study's design was not ideal, she argues that she and her collaborators were doing their best given the operational realities of police work. One problem is manpower. If another person replaces the investigating officer to coordinate a line-up, then that extra person has to be taken off

their other duties. Time and complexity are also issues. For the 40% of crimes in the Illinois Pilot Program that involved more than one perpetrator, the police found the sequential technique too cumbersome and abandoned it halfway through the study. Furthermore, real-world witnesses are often reluctant to cooperate. Their willingness to participate can hinge on the investigating officer's ability to build trust — a relationship that could be broken if the witness suddenly had to deal with a stranger conducting the line-up<sup>1</sup>. That's one reason that modern 'line-ups' are often taken to the witnesses and conducted in the field, by showing them arrays of photos.

### Ongoing dispute

Despite these difficulties, however, Mecklenburg claims that her results are still valid — even though another field study<sup>3</sup>, conducted in Minnesota at around the same time, supported the lab data. And so the dispute rumbles on.

Everyone agrees that more field studies are needed, and several are under way in the United States, including one commissioned by the Department of Justice. In the meantime, some support for the beleaguered Illinois study has come from an unexpected quarter — Britain, which prides itself on its eyewitness identification procedures. These are now routinely done in specialized suites,



**Not guilty:** Jerry Miller spent more than 24 years in jail for a crime he didn't commit.

S. STRAZZANTE/CHICAGO TRIBUNE/MCT/NEWS.COM

in the form of video parades. Video images of the suspect and foils are presented to the witness sequentially, but witnesses are asked not to give a decision until they have seen the whole sequence twice.

In a study published last year<sup>3</sup>, psychologist Tim Valentine of Goldsmith's College at the University of London, UK, and his colleagues applied strict double-blind sequential rules to British-style video parades. This was a lab study, but with what psychologists call high ecological validity — a measure of how well the experimental conditions match those in the real world. Witnesses saw a live, staged theft and attempted to identify the perpetrators a week later from a video parade constructed by police from their database.

Valentine's most striking finding was that the number of correct identifications dropped from 65% under British rules, to 36% under the strict rules. "So there is a big cost in terms of sensitivity," he says. Under strict rules, "people are inhibited when it comes to choosing". The Illinois study also found lower sensitivity, but it was overlooked in all the furore about the experimental design. If the effect is real, the implication is that the proposed reforms might reduce the wrongful conviction rate — the number of false positives, or type I errors — but at the cost of increasing false negatives, or type II errors. In other words, more criminals might be let off the hook.

### Uncertain benefits

Valentine wholly supports the introduction of double-blind line-ups, but he has serious reservations about the sequential technique. Although his group found a reduction in false positives with sequential line-ups, it was not statistically significant, at least in Britain. "In other words, we are taking a cost for a benefit we can't really be sure about," he says. Saloom disagrees, arguing that a slightly different picture has emerged from US research: for every three wrongful identifications of innocent people avoided with the sequential technique, you might let one guilty party go free. The Innocence Project understands the reluctance of law enforcers to lose that one correct identification. However, Saloom says, "Ultimately we believe that it's smarter, and a better way to protect the public, to make that larger trade-off."

Identity parades of one form or another are used in many countries. But Britain has benefited from centralized coordination of

police practices through the government's Home Office, and is regarded by many countries as advanced in the way it extracts identifications from eyewitnesses. This is in large part due to its history. In 1976, a government inquiry chaired by high-court judge Lord Devlin combed British case law, searching for wrongful convictions based on mistaken identity<sup>4</sup>. DNA evidence wasn't available then, but the committee highlighted cases in which

the crime, because there is no need to marshal a suspect or foils.

The use of DNA in a forensic context was a British invention, and it is used in a wider range of crimes in Britain than in the United States — not just serious crimes such as rape and murder. Perhaps surprisingly, therefore, no one has systematically collected data on DNA exonerations in Britain. But if they had, Wright believes, those data would show that the British system, too, is far from perfect. Now that parades take place in specialized facilities with trained officers, they are sometimes done in a blinded fashion. But the practice is still not compulsory in the United Kingdom. A spokesman from the Home Office says that before blind parades would be incorporated into the Home Office's codes of practice, they "would require evidence to support the need for change at operational level".

Indeed, perhaps the one clear lesson from this controversy is that no system is perfect. As Mecklenburg has pointed out, any reforms that psychologists propose must be workable in the real world — and they won't be unless police are given the resources and incentives required to make them work. But as Doyle points out, operational difficulties are no excuse for rejecting a fairer system. Nor are they an excuse for letting policy be swayed by a single, flawed study such as the Illinois Pilot Program. Moreover, he says, this tension urgently needs to be resolved, because the real problem isn't the study. It is the deeper clash of scientific and law-enforcement cultures. "This is really the first of the social science by-products of the DNA exoneration cases," he says. "So there is a battle being fought now about whether science is going to be allowed to affect

practices." Although mistaken eyewitness testimony is a major cause of wrongful convictions, it isn't the only one. Others include false confessions, the use of convicted informers and flawed or fraudulent scientific evidence. If the scientists back down now, Doyle warns, they will be setting a dangerous precedent. ■  
**Laura Spinney is a freelance writer based in London and Paris.**



Sheri Mecklenburg (right) and her colleague Gabriela Monahan are working to reduce Chicago's backlog of unanalysed DNA samples.

other evidence undermined witnesses' identification. The recommendations made by the committee were strongly influenced by the scientific data. For example, they picked up on the need to tell witnesses that the perpetrator might not be in the line-up.

Perhaps the most important outcome of the Devlin report, however, was to encourage a climate of cooperation between police and scientists. "I go and talk to cops and they are very interested in what I have to say," says Wright. This spirit, in turn, has led to an ongoing process of consultation and reform — the most significant in recent years being the shift from live line-ups to video parades in 2004. Among the benefits this reform has brought is that the parade can take place sooner after

1. Mecklenburg, S. H. *Report to the Legislature of the State of Illinois: The Illinois Pilot Program on Sequential Double-Blind Identification Procedures* (Illinois State Police, Illinois, 2006).
2. Schachter, D. L. *et al. Law Hum. Behav.* **32**, 3–5 (2008).
3. Valentine, T., Darling, S. & Memon, A. *Appl. Cogn. Psychol.* **21**, 933–949 (2007).
4. Devlin, P. *Report to the Secretary of State for the Home Department on the Departmental Committee on Evidence of Identification in Criminal Case* (HMSO, London, 1976).





Some researchers think that the evolution of languages can be understood by treating them like genomes — but many linguists don't want to hear about it. **Emma Marris** reports.

Consider a word as it tumbles through history: *khun*, a Nepali word for blood. In the early twentieth century, the word fell all too often from the lips of the Gurkhas, a Nepalese brigade in the British Army, in songs describing the horror of the First World War. To linguist Ralph Lilley Turner, who fought beside the Gurkhas by the Suez Canal, *khun* was one of many words transcribed phonetically for the English–Nepali dictionary he compiled in the midst of the fighting.

And then, in the 1960s, the word *khun* stopped being a word so much as a node of data in the work of linguist Isidore Dyen. Working at Yale University in New Haven, Connecticut, Dyen used Turner's dictionary to assemble a list of 200 basic Nepali word meanings, including blood, and coded them onto IBM punch cards that were fed into early computers. He used this, and similar lists from another 83 languages, to try to measure the rate at which languages change over time. His discipline of lexicostatistics is now discredited for the crude assumptions it made. But as Dyen worked on his monograph, he transferred his data on to computer disk and, in the late 1990s, published his data online<sup>1</sup>.

Nearly a century after it was sung by the

Gurkhas, *khun* became a few figures of code in the computer models of evolutionary biologist Mark Pagel at the University of Reading, UK. Last year, Pagel used Dyen's online data to generate trees that estimate when languages such as Nepali diverged from one another<sup>2</sup>. In these models, the word is stripped of much of its rich human history. But it is exactly this type of pared-down language that speaks to researchers such as Pagel the most. "Some linguists have spent a career studying a language that becomes a single data point in one of these analyses," he says. "We do things because they are mathematically elegant, and are delighted when they can be simplified."

#### A new approach

In the past five to ten years, more and more non-linguists such as Pagel have used the computational tools with which they model evolution to take a crack at languages. And one can see why. Like biological species, languages slowly change and sometimes split over time. Darwin's Galapagos finches evolved either large beaks or small; Latin *amor* became French *amour* and Italian *amore*. Darwin himself noted the 'curious parallel' between the evolution of languages and species in *The Descent of Man, and Selection in Relation to Sex*.

**"We try to find mathematical patterns in nature."**

— Martin Nowak

The advent of molecular genetics provided a new depth to the analogy. Just as the four nucleotides of DNA can produce a staggering variety of creatures, the alphabets of the world's languages can generate an infinite number of sentences. These alphabets, the words they make, and the sounds and grammar rules that frame them are passed down from parent to child in a process that, at least superficially, resembles the inheritance of DNA.

Even some complications are the same. Just as species can shade off into a maddening continuum of subspecies, populations and hybrids, languages dissolve into an untidy collection of dialects and intermediate forms. And the rampant borrowing of words between languages resembles, graphically at least, the promiscuous horizontal gene transfer that microbes engage in.

There are limits to the analogy. It is unclear, for example, what the 'selection pressures' are for language, if any. A language with a greater



number of speakers is not obviously 'more fit' than a dying language. Although a speaker of the prevalent tongue could communicate with more people, it is not the intrinsic properties of a language that make it more widely spoken. Instead, languages seem to rise to prominence on the coat-tails of the culture that speaks them, just as the prevalence of English traces back to the broad reach of British colonialism. It's no wonder, then, that mathematical biologists such as Pagel have become interested in a system that is intriguingly like, and intriguingly unlike, genes. "I think sophisticated mathematics will increasingly become part and parcel of what we mean when we say that we have 'explained' the phenomena of language change over time," says Erez Lieberman, who studies mathematics and biology at Harvard University in Cambridge, Massachusetts.

### The old school

However, there is already an old and venerable field of language-tree makers. Historical linguists have been reconstructing languages since the 1780s. Their tool is called the comparative method and it relies on extensive knowledge of the language group at hand, along with a broad grasp of, and intuitive feel for, the ways in which languages change. A linguist might notice that the way a vowel is spoken has shifted in two languages when compared with an ancient one, and infer that the shift happened before the two languages split. This will help to place the split relative to other splits but gives no information about when it happened. Hence the comparative method produces trees, but no dates.

It is putting it mildly to say that many historical linguists find the evolutionary biologists working on language histories to be bungling interlopers who have no idea how to handle linguistic data. It is also an understatement to say that some of these interlopers feel that their critics are hidebound traditionalists working on a hopelessly unverifiable system of hunches, received wisdom and personal taste. And that's just the mood between the historical linguists and the newcomers. Lots of the newcomers don't like each other either. "Why get excited about it when it is still so preliminary?" says Johanna

Nichols, a historical linguist at the University of California, Berkeley. "We are not impressed by a computational or mathematical paper per se. We have to see that it blends well with what is known by historical linguistics and really adds to our knowledge. Then we will be excited."

Perhaps the most famous and controversial study<sup>3</sup> produced by the new school is a 2003 paper by Russell Gray and Quentin Atkinson at the University of Auckland, New Zealand. The pair started with Dyen's lists of word meanings for 84 languages from the Indian and European subcontinents, plus a few extras from extinct tongues. The data already included Dyen's opinion on which of these words were 'cognates', descended from a common word in a mother language, but the researchers converted this information into numerical code and generated trees showing how and when the languages were most likely to have branched off from one another. This same type of likelihood algorithm is used to compare species' DNA sequences and produce evolutionary trees. Specifically, Gray and Atkinson dated the origin of a language family called Indo-European to around 7,800–9,800 years ago. This ineffable date has been one of the most intensely studied and disputed points in all of historical linguistics and, based on archaeological and linguistic data, had previously been put at anywhere between about 6,000 and 10,000 years ago. When Gray and Atkinson's paper made the rounds of linguistics departments, howls of protest ensued.

Some critics took the paper as a return to glottochronology, a discredited method from the middle of the twentieth century — and cousin of Dyen's lexicostatistics — which in most cases disastrously assumed that all languages change at a constant rate and which helped turn linguists against any quantitative analysis of their treasured subject. But Gray and Atkinson's statistical method does not assume uniform rates of change. Many historical linguists also felt that similarities between words are a terrible proxy for similarities between languages. They tend to argue that common sounds and grammatical rules are stronger evidence for common descent than individual words, which may be similar due to chance,

borrowing, or even 'nursery formations' such as mama and dada — words that mirror each other simply because all infants babble similar things.

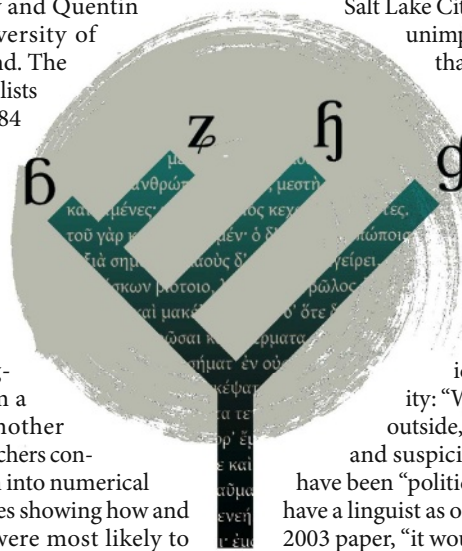
"I think that some of these researchers think that these analyses are going to supplement or even supplant historical linguists," says Lyle Campbell, a linguist at the University of Utah in Salt Lake City who was one of those unimpressed. "So far, the ones that try to go beyond what we've done don't seem to work."

Gray says that the tree does work even though it doesn't take into account the subtleties of sounds and grammar, and he puts much of the criticism down to territoriality: "When people come from outside, you see a bit of hostility and suspicion." Although it might have been "politically more palatable" to have a linguist as one of the authors on the 2003 paper, "it wouldn't have changed the answers", he says.

Ultimately, many linguists felt that this type of analysis oversimplified their cherished subject more than they could bear. Linguists love the little details that give a language personality: to them, the identifying sounds or peculiar borrowed words are nuances that tell the tale of a tongue. The new breed brushes over these details in pursuit of generalities, trends and statistical rules. "We try to find mathematical patterns in nature," says Martin Nowak, an evolutionary modeller at Harvard. "If someone works on the detailed classification, they might be dissatisfied with something that is cruder."

### Grand ambitions

That dissatisfaction looks set to grow as many in the new school pursue grander ambitions: to find quantitative laws that describe language evolution. In a recent example, published in *Science* earlier this year, Pagel, Atkinson and their colleagues used word lists to build trees in three of the world's major language families; Indo-European, Bantu (an African language family that includes Swahili) and Austro-nesian — spoken on Pacific Islands<sup>4</sup>. They found that between 10% and 33% of divergence among these languages happened in what they called 'punctuational bursts', phases



**"We do things because they are mathematically elegant, and are delighted when they can be simplified."**

— Mark Pagel

of accelerated language evolution just after a language splitting event. The finding echoed the controversial 'punctuated equilibrium' theory in which Niles Eldredge and Stephen Jay Gould proposed that biological evolution often occurred in rapid bursts amid longer periods of relatively slow change. Pagel and his team speculate that the bursts could arise from the spoken idiosyncrasies of a small number of population founders, or a desire within a new population to sound different from the other group. So here is one general law, perhaps: up to one-third of language evolution occurs in punctuational bursts after splitting events.

A second possible law arose from studies of word frequency. In their 2007 study, which was published in *Nature*, Pagel and his team found that 50% of the difference in language evolution rates could be explained by the frequency with which words within the language were being used<sup>2</sup>. Often-used words were 'stickier' and resisted change. "What really excites me about the frequency effect is that we are identifying a general evolutionary law," Pagel says. "We think it will hold and will have held since we began talking."

In the same issue of *Nature*, Lieberman, Nowak and their co-workers showed that irregular English verbs become regularized more quickly if they are rarely used<sup>5</sup>. So the past tense of a rare verb such as 'gnaw' would have a 50% chance of regularizing to 'gnawed' from the Old English form 'gnagan' in 700 years. By contrast, a very common verb such as 'be' would have a 50% chance of regularizing to 'beed' in 38,800 years, perhaps explaining why 'was' remains the preferred form today. The researchers even had a precise mathematical description of the trend: a verb that is used 100 times less frequently regularizes 10 times as fast.

### Different language

These findings completely underwhelmed most historical linguists. They already knew that commonly used words change more slowly, and the fact that some aspects of this trend could be quantified did not really interest them. "I don't think the numbers are very exciting," says Campbell. "I would much rather it be relativized to 'in general, more frequent words change more slowly'"

One reason that many linguists have been

unreceptive to such work is that they are not trained in statistics, and are unsure of how to compare and evaluate this type of numerical model themselves. "They feel they've been asked to just accept things," says Tandy Warnow, a computer scientist at the University of Texas at Austin who works with linguists. And even those people, such as Warnow, who can evaluate the models say that they are too unsophisticated at this stage to pronounce firm dates or quantitative rules. She says that the biological models need to be tailored to language and that they should incorporate the sound and grammar changes that are so important to linguists.

### Better representations

Paul Heggarty, a linguist at the McDonald Institute for Archaeological Research at the University of Cambridge, UK, is already trying to refine his models in this way. Heggarty is building network diagrams rather than trees to show how similar languages interrelate. He thinks that these can provide a better representation of the relationship between two languages when two cultures rub shoulders very closely and borrow words freely. Then the links between branches of the tree — equivalent to horizontal gene transfer — become more important than the vertical branching. "It is entirely natural for languages to stand in complex cross-cutting relationships to each other that may not be compatible with any branching genealogy at all," he says.

As part of the network building, Heggarty is also trying to assign more subtle values to

word changes than zeroes and ones.

A superficial analysis of the word 'dog', for example, might show that the English word is not cognate to the German word (*hund*) and score 1, or 'changed' for the pair. But if the English word 'hound' is chosen instead, it creates a match and would score 0. Because 'hound' isn't the main word for 'dog' in English, Heggarty would score it somewhere in between 0 and 1, perhaps 0.4. He hopes that this type of refined method can create networks that reproduce the real relationships between very closely related languages and, by extension, reveal something about the histories of the peoples who spoke them in the past.

### Getting quantitative

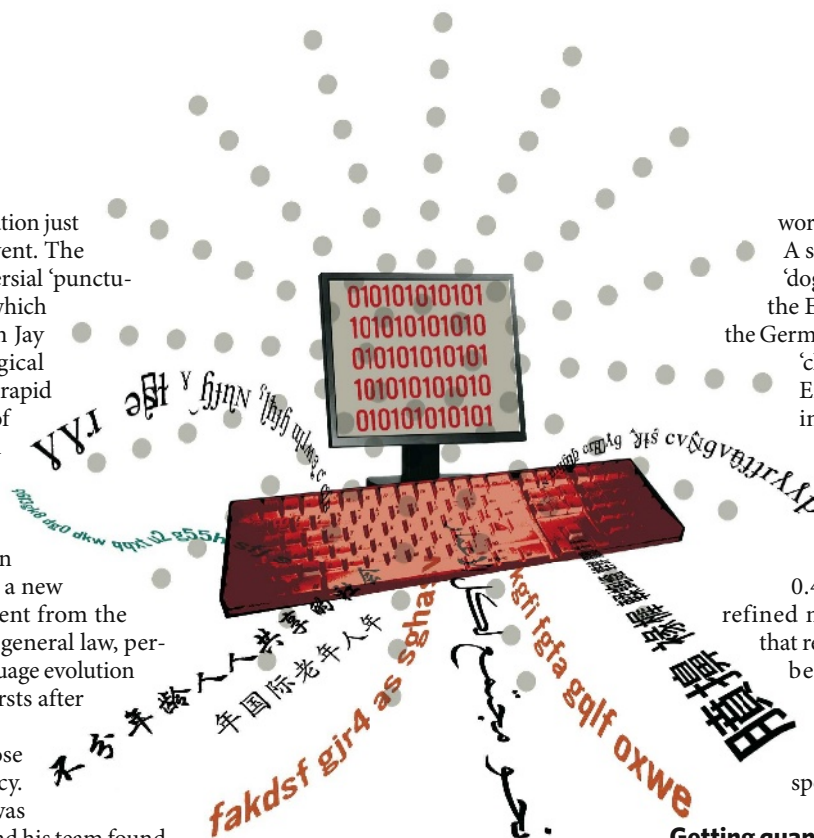
Such model tweaking is unlikely to win over the historical linguists, but at least some are beginning to warm to the methods. Campbell acknowledges that the sheer number-crunching power of computer models can speed up the good old comparative method. And he sees the appeal in getting a bit more quantitative. If the field does not become more statistical and accountable, he points out, it may lose respect by those in other disciplines. "I think we'd like the legitimacy," he says.

Another fan is Harvard University's Steven Pinker, who famously appreciates language in all its fullness. "There has got to be information in the statistics of language overlap that you simply can't exploit by looking at it intuitively, by eyeballing," he says. "Linguists have been slow in accepting that extra dollop of information that statistics provides, even if there are errors, even if there is noise."

Noise — of the statistical kind — is not comfortable territory for many historical linguists when precious words such as *khun* are at stake. So perhaps the onus now lies on the newcomers to show that their methods will not drown out languages, or their rich and idiosyncratic narrative. "Hope," Pinker says, "is not that the older generation of linguists will lay down their arms; hope is that the younger generation will follow their noses to what is fruitful."

**Emma Marris writes for *Nature* from Columbia, Missouri.**

1. Dye, I., Kruskal, J. B. & Black, P. *Trans. Am. Phil. Soc.* **82**, 1-132 (1992).
2. Pagel, M. *et al. Nature* **449**, 717-720 (2007).
3. Gray, R. D. & Atkinson, Q. *Nature* **426**, 435-439 (2003).
4. Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J. & Pagel, M. *Science* **319**, 588 (2008).
5. Lieberman, E. *et al. Nature* **449**, 713-716 (2007).





## CORRESPONDENCE

## Acceptance of peer review will free Italy's research slaves

SIR — 'Italy must invest more in science and technology' according to I. Bertini, S. Garattini and R. Rappuoli in Correspondence (*Nature* **452**, 685; 2008). They lament the Italian lack of financial resources and political attention for research, technology and education. As a researcher, clinician and academician, I share their concerns. However, as former chair of the health committee of the Italian Senate, I take exception to their implication that none of the major political parties recognizes science, technology and education as crucial for the future of the country's economy.

The 2007 and 2008 national budget laws, drawn up when the centre-left coalition was in power, allocated €96 million (US\$149 million) to projects submitted by researchers under 40 years old. These are judged by an international committee comprising ten scientists under 40 — five from foreign institutions — selected according to impact factor and citation index scores. This alone is a revolutionary approach for the unregulated Italian system of research funding allocation.

In spite of such advances, Italy is still far behind in research investment, and this needs to change. But the crucial switch is not simply to increase funding. The way the new government should proceed is to reform the allocation criteria for funding and to start applying across the board the selection and evaluation rules of peer review. Such a system would acknowledge meritocracy and free researchers from the virtual slavery under which they have been kept by old academicians.

By applying international rules of peer review and evaluating grant applications only on the basis of merit, looking at curricula and objectives,

comparing lists of publications and evaluating results, we will provide opportunities for Italy's scientists, thereby promoting the country's intellectual, cultural and economic growth.

**Ignazio R. Marino** Department of Surgery, Jefferson Medical College, 19107 Philadelphia, Pennsylvania, USA, and Senate of the Republic of Italy, Piazza Madama snc, 00186 Rome, Italy

## Mimicking photosynthesis, but just the best bits

SIR — Your News Feature 'The photon trap' (*Nature* **452**, 400–402; 2008) makes good points about the challenges for converting solar to fuel energy by artificial photosynthesis.

But we wish to clarify the assessment that "simply mimicking photosynthesis is too short-sighted".

The (highly optimistic) 3% efficiency for solar energy conversion in plants covers everything that a plant gets up to, day and night, during an annual cycle. The whole complex process of photosynthesis, not to mention the plant's way of life, is certainly not a target for chemical mimicry. Biologically inspired chemistry based on photosynthesis focuses only on the specific reactions that are potentially useful.

Early aviation pioneers, who looked to birds for biomimetic aeroplane design features, incorporated wings, a tail, a fuselage and aspects of aerodynamics into their final product. In the main, they chose not to go for flapping — and nest-building and flying south for the winter were right out. Biomimetic chemistry is the same: we pick only the relevant bits.

The water-oxidizing enzyme

you feature is currently the focus of attention — a chemical marvel in which a low-energy pathway removes electrons from water so that the enzyme can operate at minimum electrical over-potential. Its high energy-conversion efficiency is unmatched by artificial catalysts derived from cheap and abundant elements.

Your News Feature implies that research on this enzyme has advanced to the point where it can provide a legitimate target for catalyst-hunting chemists to mimic. This is true, but there is still a good deal to be learned about the structure and mechanism of the enzyme itself and this will doubtless be of great benefit to future research on artificial photosynthesis.

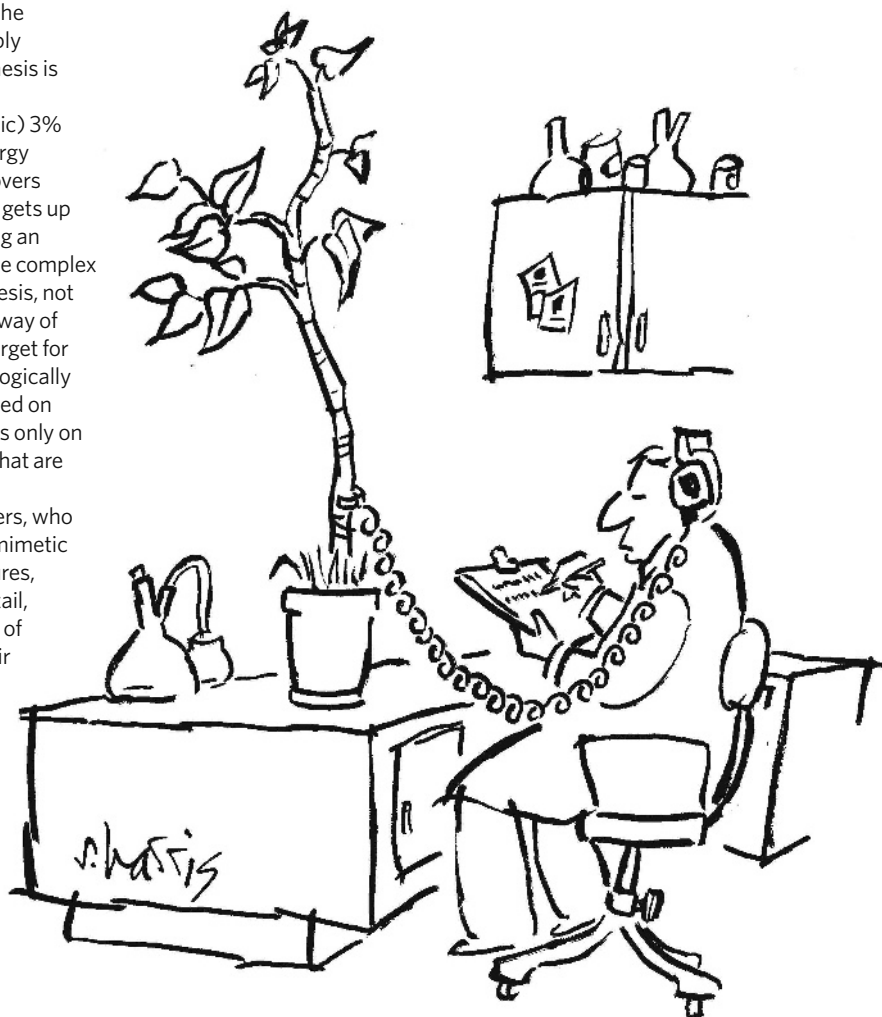
**A. William Rutherford** iBiTEC-S, URA

2096 CNRS, CEA Saclay, 91191 Gif-sur-Yvette, France

**Thomas A. Moore** Center for Bioenergy and Photosynthesis, Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA

## Standard identifier could mobilize data and free time

SIR — The rise of bioinformatics has focused attention on the growing depth and scope of database content. However, it is difficult or impossible given the existing citation metrics system to identify who originally created or added value to a datum. Without a system to reward, we shall continue to rely on the good will





or spare time (sic) of researchers to mobilize data into the public domain.

One proposal discussed recently (<http://tinyurl.com/6elpq4>) concerned the building of a realistic measure for use of database elements and a 'cite me' button for a dynamic composite web page. The Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)) is investigating the assignment of 'life-science identifiers' to allow accreditation not only to data sets, but also to the individual datum and its author. Simplified mechanisms are needed that make it easy for individuals to assign these identifiers to their data.

We believe that scientists' productivity also needs to be gauged through data publishing, which requires a culture change in the recognition of scientific output. An industry-standard identifier, such as that proposed by the GBIF, could be part of publishers' referencing systems, and authors could provide 'citation identifiers' for all data records and data sets. Such a mechanism would achieve increased data mobilization and increased accreditation, both desirable to scientists.

**Dave Roberts** The Natural History Museum, Cromwell Road, London SW7 5BD, UK

**Vishwas Chavan** GBIF Secretariat, Universitetsparken 15, DK-2100 Copenhagen, Denmark

## Name variations can hit citation rankings

SIR — The Correspondence 'Give south Indian authors their true names' (*Nature* **452**, 530; 2008) and earlier News Feature 'Identity crisis' (*Nature* **451**, 766–767; 2008) are highly relevant to calculations of PubMed citations and h-index (the number  $n$  of a researcher's papers that have received at least  $n$  citations).

For example, I used to use the south Indian form of my name:

T. Biji Kurien, with Biji being my personal name. I have seven publications cited incorrectly in PubMed as being by 'Kurien, T. B.', 'Bijikurien, T.' or 'Kurien, B.'. Four of these entries were cited often enough to be counted towards my h-index computation. As I had by then changed my name to conform with Western style, these publications unfortunately do not appear in the Web of Science or PubMed under my current name format. Consequently, my h-index ranking has fallen by 25%.

It is of paramount importance to adhere to a consistent name pattern right from the start, in order to maintain a correct list of publications in the public databases as well as the right h-index rankings.

**Biji T. Kurien** Arthritis and Immunology, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, Oklahoma 73104, USA

## Names: dropped to avoid prejudice, now useful again

SIR — The Correspondence 'Give south Indian authors their true names' (*Nature* **452**, 530; 2008), incorrectly states that people from the south do not traditionally have surnames.

I am from southern India and have a proper surname — as do all the families in my region. Besides Patil, surnames such as Naidu, Reddy, Rao and Gouda are common in the different states of southern India. One of the authors of the Correspondence has the surname Kutty.

Surnames have widely fallen into disuse because our fathers and forefathers avoided using them to prevent discrimination on grounds of caste.

It doesn't make sense in this case to use only an author's first name in scientific publications and to devise a special system to accommodate a different naming format. Instead, editors should

encourage these authors to revive the use of their surnames.

**Prabhu B. Patil** E307, Centre for Cellular and Molecular Biology, Hyderabad 500007, India

*Readers are welcome to comment at the Nature India blog Indigenus, <http://tinyurl.com/58r9wf>*

## Open-access more harm than good in developing world

SIR — The traditional 'publish for free and pay to read' business model adopted by publishers of academic journals can lead to disparity in access to scholarly literature, exacerbated by rising journal costs and shrinking library budgets. However, although the 'pay to publish and read for free' business model of open-access publishing has helped to create a level playing field for readers, it does more harm than good in the developing world.

Authors by no means have a level playing field, even in the traditional publishing model. The dynamics of peer review make it hard to ensure that publication of an article is a function of only its quality, uninfluenced by factors such as topicality or the author's name and affiliation. The open-access model makes the playing field for authors even more uneven.

Page charges may be waived for authors who cannot afford to pay, but a model that depends on payment by authors can afford only a few such waivers. And why should anyone want to survive on charity? The argument that it is the granting agency and not the author that pays does not wash either. If anything, the playing field for grants is even more uneven. Besides, this will undermine, rather than encourage, the whole area of grant-free research.

Page charges make extra difficulties for authors, while the old problems associated with peer review persist. They could be disastrous for the underdeveloped

world, encouraging people to remain as consumers (readers), rather than to become producers (authors) of knowledge.

A 'publish for free, read for free' model may one day prove to be viable. Meanwhile, if I have to choose between the two evils, I prefer the 'publish for free and pay to read' model over the 'pay to publish and read for free' one. Because if I must choose between publishing or reading, I would choose to publish. Who would not? **Raghavendra Gadagkar** Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India

## A 3D revolution in communicating science

SIR — Since the release of Adobe Systems' Portable Document Format (PDF) version 1.6 in 2004, it has become possible to view interactively three-dimensional models that are embedded into PDF files. This attribute will dramatically increase information content as well as data transparency in scientific papers. Additionally, replacing multiple two-dimensional figures of a three-dimensional structure with one integrated interactive three-dimensional model will reduce the need for supplementary material.

The potential of this technological advance for all science is obvious. Because of the foreseeable rise in demand by the scientific community, publishers and scientific institutions need to work hand in hand to support the implementation of this highly desirable technique.

**Jérôme Murienne** UMR 5202, Département Systématique et Evolution, case 50, Muséum national d'Histoire naturelle, 45 rue Buffon, 75005 Paris, France

**Alexander Ziegler** Institut für Biologie, Freie Universität Berlin, Königin-Luise-Straße 1-3, 14195 Berlin, Germany **Bernhard Ruthensteiner** Zoologische Staatssammlung München, Münchhausenstraße 21, 81247 München, Germany

## BOOKS &amp; ARTS

# Security in an uncertain world

Biological protection systems that have evolved over billions of years could be the key to strengthening national defences against unforeseen threats, says **Jessica Flack**.

## Natural Security: A Darwinian Approach to a Dangerous World

Edited by Raphael D. Sagarin and Terence Taylor

University of California Press: 2008.  
289 pp. \$49.95, £29.95

In 1957, commenting on the power balance between the Soviet Union and the United States, physicist Robert Oppenheimer said: “In time, the transnational communities in our culture will begin to play a prominent part in the political structure of the world, and will even affect the exercise of power by the states.” Writing in 1986 in *The Making of the Atomic Bomb*, Richard Rhodes interpreted Oppenheimer’s transnational community as that of science, arguing that with the invention of the atomic bomb, “science became the first living organic structure strong enough to challenge the nation-state itself”.

Since the end of the cold war, during which relative stability prevailed, threats to national security have become unpredictable. Oppenheimer’s comment foreshadowed the growing role of science, particularly physics, in international politics. It also foreshadowed the current source of the unpredictability: loosely organized, transnational networks of individuals seeking to attack nation-states.

In this uncertain age, we might look to an evolutionary theory of organizational robustness to provide a basis for a predictive science of national security. A good starting point is the engaging book *Natural Security*, edited by ecologist Raphael Sagarin and security expert Terence Taylor. Political scientists, anthropologists, ecologists, epidemiologists, evolutionary biologists and palaeontologists share lessons from 3.5 billion years of experimentation by biological systems in maintaining their security in a hostile and unpredictable world.

The concept is not new. For thousands of years, humans have sampled nature’s strategies to improve their quality of life. What is new is the idea that by studying how organisms survive unpredictable events, we might identify general principles that apply to national security. Sagarin introduces the book by identifying critical questions: when do major shifts occur in human and natural systems? What types of organisms survive mass extinctions? And which events lead to



S. HUNT/GETTY IMAGES

The porcupine fish evolved spines to protect it from attack in its aquatic environment.

escalations of armaments and defences?

Rather than being built around these foundational questions, *Natural Security* is organized around scientific disciplines. The book does not offer an analysis of principles but a diverse sampling of potential solutions to problems of national security drawn from observing the history of life. A danger of this approach is that solutions that seem to be generic are not, having evolved in a particular context and with a particular set of supporting mechanisms. In addition, as Sagarin and evolutionary biologist Geerat Vermeij note, nature can experiment without ethical concern for study subjects and risks arising from failure, whereas societies cannot.

The book would have been more compelling had it advocated a systematic study of what works and why, and at what cost. It might have been organized around the three main classes of robustness mechanisms observed in stable systems in the biological world — management, repair and prevention.

Management mechanisms control the spread and severity of damage induced by perturbations, either by actively countering them

or by using structural tactics that maintain functionality despite damage. Virologist Luis Villarreal explains how humans have three immune systems to block attacks. The innate immune system builds barriers such as skin to keep pathogens out; the adaptive immune system can recognize, respond to and improve its response to invading foreign agents; and a ‘behavioural immune system’ excludes infected individuals socially. The book might have explored the implications of adopting a multi-tiered defence system for homeland security, with mechanisms operating on different timescales and tuned to different kinds of perturbations.

Repair mechanisms allow a system to rapidly recover its initial state. Ferenc Jordán, an ecologist who studies food webs, suggests that stability can be increased by building networks with links that can be rewired to maintain connectivity if parts of the network are damaged. Analogously, disaster-relief systems could establish back-up relationships among relief agencies to ensure that bottlenecks do not hinder the distribution of emergency resources.

Preventative mechanisms can reduce the likelihood of perturbations by altering the environment to reduce conflicts of interest between parties, or to create dependencies that are beneficial. One explanation for the evolution of the arrest of meiosis, the process by which gametes are produced, is that early sequestering of the germline protects it by minimizing the total number of possible mutations. In this way, conflict is pre-emptively eliminated. Bradley Thayer, an expert in national security, suggests that the motivation behind the US policy of spreading 'effective democracy' is to change the environment from one that fosters extreme positions to one that is open to negotiation. By drawing on analogous processes in biology, one might

show the conditions under which such policies are likely to work.

Robustness has its costs. The trade-off between robustness and the ability of a system to reconfigure into a new state when faced with a changed environment — known as evolvability — is poorly understood in evolutionary theory. The consequences for the evolvability of the mechanisms discussed in *Natural Security* are unknown, and these ideas should be adopted with caution. Modularity, for example, may allow reconfiguration and limit damage by decoupling the fates of components and providing a flexible architecture. However, coordinating the different parts can be costly and difficult to manage. In hunter-gatherer societies, the division of

labour requires the building of a distribution system supported by exchange rules; if the rules are unclear or violated, then conflict can result. When components are too specialized, their ability to adopt other functions is sometimes lost, making the system less evolvable and less robust.

*Natural Security* is a stimulating read. It opens the door to an exciting merger between political science and evolutionary theory. The task now is to use the ideas of organizational robustness that are developing in evolutionary theory to formulate principled hypotheses about the consequences of national-security decisions. ■

**Jessica Flack** is a research fellow at the Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

## Genetic medicine at the bedside

### **Heredity and Hope: The Case for Genetic Screening**

by Ruth Schwartz Cowan

Harvard University Press: 2008. 270 pp.  
\$27.95, £18.95

Despite the fresh veneer of technology, medical genetics still follows the old-fashioned practice of medicine. It remains the most clinical of disciplines — in the literal sense, from the Greek *klinikos*, meaning 'of a bed' — in that most of the genetic physician's work is done at the bedside.

The story of the patient's illness, their family history and the physical examination remain the cornerstones of diagnosis. A clinician must examine the whole body to catalogue subtle and obvious signs and symptoms: the texture of the skin, how the ears are slung, the shape of the uvula in the back of the throat. Clinical findings then cohere, much like stars in constellations, into the eponymous syndromes with which we are familiar.

In *Heredity and Hope*, technology sociologist and historian Ruth Schwartz Cowan writes brief histories of several hereditary diseases and the scientists and clinicians who developed screening tests for them. Of the thousands of genetic diseases, Cowan focuses on a handful that are atypical in that they are well understood biochemically, genetically and sociologically.

These include Tay-Sachs disease and phenylketonuria, which result from enzyme deficiencies, and sickle-cell anaemia and  $\beta$ -thalassaemia, which arise from defects in  $\beta$ -haemoglobin, one of the most studied of all proteins. For each disease, the probability of clinical expression given a specific genotype is very high, making predictions reliable and early detection routine.

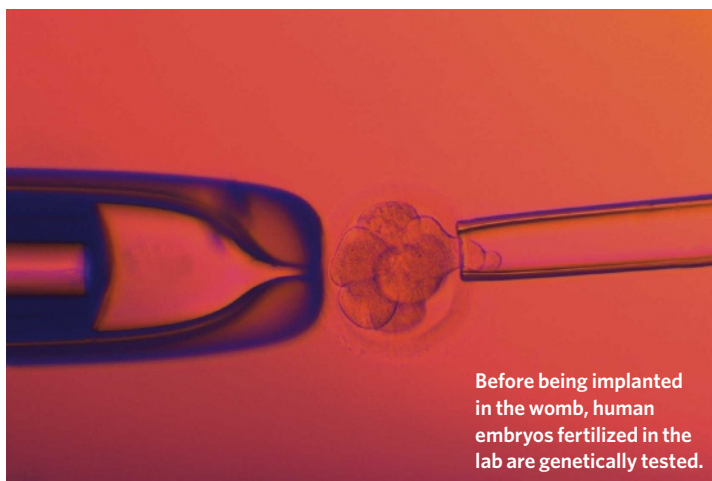
The consequences of these diseases remain devastating to patients and their families. This is especially true in the case of phenylketonuria, where a delayed diagnosis may result in irreversible brain damage. Physicians, parents, patients and insurance providers all agree on the benefits of identifying carriers of the mutant genes or diagnosing disease either *in utero* or at the time of birth, and identification protocols have been crafted that are acceptable to most. The greatest disagreements

centre on what action to take once we have this genetic information.

The author's brief history of eugenics presses the point that medical genetics owes no apologies to society. There is no overlap between those who care for patients with genetic disease and anyone who has advocated the purification of the general germplasm through genetic isolation, including sterilization. This is obvious given that eugenics as public policy and as science met its deserved end in the first half of the twentieth century, whereas medical genetics as a sub-speciality formally began in the 1950s when Victor McKusick opened the Moore Clinic at the Johns Hopkins Hospital in Baltimore, Maryland.

That medical genetics and eugenics sprang from the same scientific soil has given ground to a small chorus of opponents to genetic screening. Trying to pull the ugly thread of eugenics through the fabric of genetics to discredit it, these opponents range from what Cowan calls 'reproductive feminists' to advocates of rights for people with disabilities, and span both the political left and right. This is not to dismiss the defensible reasons to object to population-based screening for specific diseases.

Clinical variability can be huge, even for specific genotypes, so the decision to establish a screening programme is not straightforward. Every medical geneticist has been confronted by the fluid meaning of disability. Despite clear clinical challenges, many deaf people, for example, do not consider themselves disabled and rightly



Before being implanted in the womb, human embryos fertilized in the lab are genetically tested.

P. GOETHEL/SCIENCE PHOTO LIBRARY



resent being defined by their trait. Medical geneticists have had to adapt to patients' views of themselves.

To justify morally the genetic-screening programmes she writes about, Cowan cites the good intentions of the parties involved, primarily their efforts to relieve suffering. This criterion does not pass philosophical muster, nor is it sufficient to sway vehement opponents. Although her analysis is cursory, it does get to the heart of the matter: hereditary diseases cause great human suffering and everyone wants to help.

What no commentator on medical genetics acknowledges is the hidden sadness, customarily buried, that each geneticist feels when discussing with patients and parents the options for treatment, which are generally few and unsatisfactory. This takes its toll on everyone, although patients always astound with their resilience.

The hard truth is that genetics does not offer easy answers. There are many genetic diseases, and each one is unique. The simplicity of DNA

is illusory — our DNA is popularly regarded as our medical fate, but DNA interrogations more often yield notions of risk that have different meanings to patients and physicians. Physicians rarely know the true cause of our complaints. In those genetic cases where simplicity prevails, the testing technology is likely to be adopted. As law in the United States, the Genetics Information Non-discrimination Act may relieve some anxiety about the misuse of genetic information. If only it were so simple to dispatch misery.

Modern healers may claim science to be the foundation of their work, but the key is, in fact, persuasion: to heed advice, to push and persevere, to hope. As the genome is further dissected and better understood, no family of diseases warrants more genuine hope for successful management than genetic conditions. Cowan understands that we must all share that hope for the campaign to be successful. ■

**Hugh Young Rienhoff Jr** is director of MyDaughtersDNA.org, based in San Francisco, California, USA.

## A rough guide to Titan

### **Titan Unveiled: Saturn's Mysterious Moon Explored**

by Ralph Lorenz and Jacqueline Mitton  
Princeton University Press: 2008. 296 pp.  
\$29.95, £17.95

A future tourist guidebook to this remote destination would warn us to bring our heavy-duty rain gear, but be prepared not to need it. Droughts may last many years there, but when a hurricane-sized storm sweeps across the sky, the rainfall is torrential. At high latitudes, the landscape is dotted with thousands of lakes, some mere ponds and others inland seas. Networks of channels and canyons are etched into the terrain, over which huge volcanic domes loom. Other regions harbour vast fields of dunes, some 100 metres high. Welcome to Titan, Saturn's largest moon.

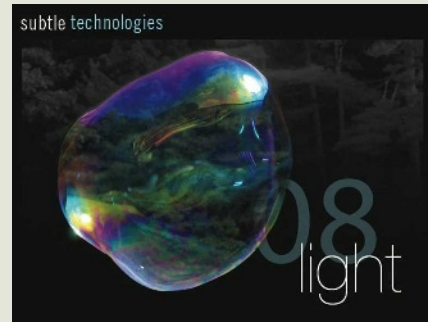
Our guidebook would go on to explain that the dune particles are not sand, but hydrocarbons, totalling more than all the coal reserves on Earth. The magma flowing from the volcanoes is not liquid rock, but a mix of ammonia and water, similar to antifreeze. Liquid ethane fills the lakes. And liquid methane carved the gullies at rates far in excess of the worst flash-flooding on Earth.

*Titan Unveiled*, by planetary scientist Ralph Lorenz and astronomy writer Jacqueline Mitton, presents a good overview of the

state of our knowledge of this curious moon, and is accessible to most. Lorenz is closely involved with the Cassini mission to Saturn and the Huygens probe it dropped onto Titan's surface in 2005. The book focuses on his key interests, which include Titan's surface and lower atmosphere, regions that parallel Earth and are thus the most engaging for readers.

*Titan Unveiled* describes how most of what we once hypothesized about Titan has been proved wrong. The story of how we gained our current knowledge is fascinating; even more intriguing is what remains to be learned. Larger than Mercury, Titan is the only moon in our Solar System that is enveloped in a thick atmosphere. Analogous to Earth's water-based weather system, Titan's atmosphere experiences weather based on the phase changes of methane, shifting between its gas, liquid and solid states. At the extremely cold temperatures on Titan's surface (94 K, or  $-179^{\circ}\text{C}$ ), water is frozen and acts like rock. The moon is geologically active, including volcanism and uplifting of mountain ranges. Deep under the icy surface, evidence for an ocean of liquid water and ammonia has been found.

Scattered throughout the text are personal anecdotes by Lorenz, labelled "Ralph's Log". Key to the book's success, these sections



WWW.SUBTLETECHNOLOGIES.COM

### **NORTHERN LIGHTS**

The many shades of light in art and science are the focus of the annual Subtle Technologies Festival in Toronto, Canada, starting this week. A symposium (from 30 May to 1 June) will discuss the physics of light, its use in education, photography, performance, new media and architecture. Sound artists muse about synaesthesia; a physicist explains why painters love the light in Provence, France; and a biologist describes how to image cells.

[www.subtletechnologies.com](http://www.subtletechnologies.com)

### **GREEN FINGERS**

Gardeners are the canaries of climate change: first to notice buds blooming early, lawns that need mowing more often and pests spreading in range as average temperatures creep up. This week's Chelsea Flower Show in London (until 24 May), run by the Royal Horticultural Society, includes scientific exhibits to educate plant lovers about climate change. UK researchers from the Tyndall Centre in Norwich, the University of Reading, Rothamsted Research in Harpenden, and others will be on hand to explain how plants, ecosystems and practices must adapt.

[www.rhs.org.uk/chelsea/2008](http://www.rhs.org.uk/chelsea/2008)

### **BEING HUMAN**

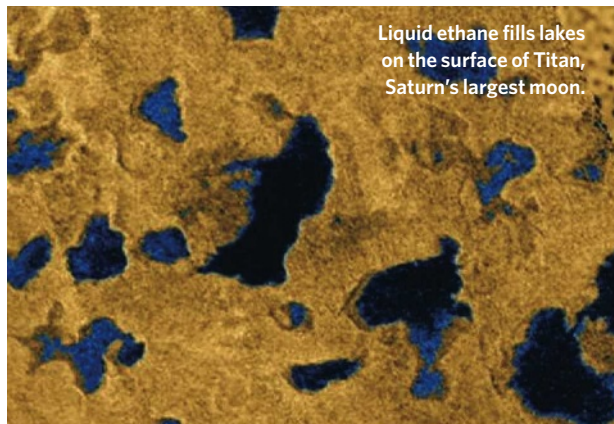
An exploration of what it means to be human in a rapidly changing world and vast Universe is the theme of the 55th Carnegie International. The largest US survey of contemporary art, it opened this month at Pittsburgh's Carnegie Museum of Art (until January 2009). *Life on Mars*, named after David Bowie's song, offers 300 works from 40 international artists, including Vija Celmins, who received the US\$10,000 Carnegie Prize for her *Night Sky* series of paintings.

[www.cmoa.org](http://www.cmoa.org)

CULTURE DISH

convey how planetary exploration, and science in general, progresses as a human enterprise. Lorenz communicates what it is like to be a scientist involved with a current space mission, working with diverse colleagues and following your curiosity to make new discoveries.

Advances may come serendipitously, but they are usually hard-won following years of intense work, carried out with the risk of failure and research dead-ends. Some obstacles to progress are simple to overcome. For example, Lorenz recounts how, while working alone at night at an observatory, he was once held back by a crucial piece of equipment that lay behind a locked storage-room door. His eventual solution was to remove the door's hinges. Other challenges are greater, such as the discovery of an engineering problem with the radio transmitter on the



Liquid ethane fills lakes on the surface of Titan, Saturn's largest moon.

Huygens probe after its launch. It required a major effort to retarget and replan nearly the entire mission, involving hundreds of people and thousands of hours of work.

With the Cassini mission flying past Titan every few weeks and astronomers observing

it from Earth nearly every night, new discoveries are regular. It is inevitable that any book on Titan is a little out-of-date before it is released, but this reflects the vitality of the research.

We won't be able to book a ticket to Titan in the next few decades, but further robotic spacecraft will be sent to explore. A Titan orbiter could map the surface, observe the seasonal weather patterns and study the subsurface ocean. Balloon-borne detectors could examine the atmosphere and surface up close. And a new mission will add detail to our guidebook to Titan. Hopefully,

someone working on that mission will write an insider's account, like *Titan Unveiled*, to tell us how it all happened.

**Henry Roe** is an astronomer at Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, Arizona 86001, USA.

AP PHOTO/NASA

## How science hit the small screen

### Films of Fact

Science Museum, London

From 29 May to 2 November 2008.

**Films of Fact: A History of Science in Documentary Films and Television** by Timothy Boon

Wallflower Press: 2008. 224 pp.

£45.00 (hbk), £16.99 (pbk)

"Is it not a scandal, in this day and age, that there seems to be no place for continuing series of programmes about science?" asked veteran natural history broadcaster David Attenborough, lecturing on the future of public service television in London on 30 April. "If you want an informed society, there has to be a basic understanding of science."

An exhibition opening next week at London's Science Museum, *Films of Fact*, charts how science was introduced to the UK public in documentary films and on television in the early twentieth century, from the birth of these media to the 1960s.

Animals and plants featured in the first science films made for public viewing. Lasting for 56 seconds, the 1903 film *Cheese Mites* was first screened at London's Alhambra Music Hall as part of a musical and theatrical playbill that included ballet and magic tricks. Filmed down a microscope by amateur natural historian Francis Martin Duncan, the greatly magnified mites scuttle about. They may not seem riveting to our jaded eyes, but

they stimulated demand for nature-based films. Producer Charles Urban exploited this commercial potential in a series of photomicrography films called 'The Unseen World: Revealing Nature's Closest Secrets by Means of the Urban-Duncan Micro-Bioscope', which included *The Circulation of the Protoplasm of the Canadian Waterweed* (1903). Nature series quickly became established as a popular genre and remain so today, from movies of meerkat antics to marching penguins.

The most successful nature film series



Comic chemistry: Rotha's 1938 *New Worlds for Old*.

before the Second World War was *Secrets of Nature* (1922–33), produced by British Instructional Films. Its successor was *Secrets of Life* (1934–50). Celebrated cameraman Percy Smith, a clerk at the UK government's Department of Education, worked on both series. He specialized in filming through microscopes or glass aquaria in his London greenhouse, using a timing device he made from a cuckoo clock to record plant growth with time-lapse photography.

Television programming about science took off in the mid-1950s in the United Kingdom, two decades after broadcasting began there in 1936. Some science series were designed to teach. Producers and scientists worked together, mostly in live broadcasts such as *Eye on Research* (1957–61), which took cameras into research establishments.

As television became a mass medium, scientists tried to influence how broadcasters represented science, but they did not always get a good reception. "Priority must be given to the medium rather than scientific pedantry," ruled Aubrey Singer, head of the BBC's science department in 1966. "The aim of scientific programming... is not necessarily the propagation of science" but "an enrichment of the audience experience". Similar attitudes prevail today.

Other documentaries, many commercially sponsored, explored how new technologies were transforming everyday life. Influential film-maker Paul Rotha's 1933 documentary *Contact*, sponsored by Imperial Airways, captures aeroplane manufacture using beguiling and original cinematography. From the

© NATIONAL GRID PLC; SUPPLIED BY THE BRITISH FILM INSTITUTE



1930s, Rotha and others used angled shots and rapid editing — techniques pioneered by Russian film directors and cinematographers — to celebrate innovations such as aircraft, telephone networks, electricity and express railways. Enthusiasm for technology remains a strong driver of scientific film and television.

The Depression of the 1930s stimulated documentaries in which scientists identified social problems and proposed solutions. In 1936, *Enough to Eat?* relayed the shocking conclusion of nutritionist John Orr, in a study entitled *Food, Health and Income* for the UK Ministry of Agriculture, that half the population of the United Kingdom was too poor to maintain a healthy diet.

In the exhibition, film and television clips are projected onto a screen, and hundreds of other clips from 38 films can be accessed interactively though two computer stations. Pieces of film-making equipment are also on show: a Moy and Bastie cine camera made to

Urban's design; a Zeiss microscope of the type used by Smith; a Marconi IV studio television camera used in the 1960s; a Moviola editing machine; and a 1930s Newman Sinclair cine camera.

Chief curator of the Science Museum, Timothy Boon, has written a well-researched book that provides background detail for historians of UK science film-making during this period. Other researchers are tackling French, Russian and US depictions of science on film and television, plundering those nations' archives with equal diligence. Once these studies are complete, it would be valuable to combine them into a global account of science on screen.

"My ambition for the show," says Boon, "is that by seeing different types of science films, people will become more informed consumers of science television now." Hopefully, greater knowledge of how science programming developed will guide decisions about its future. ■

Colin Martin is a writer based in London.

unusual elegance and beauty. A 1997 green and turquoise gown by Thierry Mugler, for example, seems to be destined for a creature that is part bird, part crustacean; long-sleeved with a flowing train, it consists almost entirely of feathers, its middle a segmented carapace. Spider-Man stirs skiwear designer Spyder, whose web-patterned race suits are on display, as well as Giorgio Armani, whose offerings include a 1990 beige evening dress sheathed in a delicate web of insect-adorned netting.

Body armour also enthralles avant-garde fashionistas. The shield of superheroes such as Iron Man — played in this spring's blockbuster by Robert Downey Jr, whose LED-eyed fibreglass costume is on show here — finds new forms in such ensembles as Gareth Pugh's 2007 leather-and-synthetic dress. With sleeves formed of shiny, triangular black panels, it resembles a solar-powered bat. Speaking of bats, the show does a nice sideline on stylish dominatrix wear, as epitomized by Michelle Pfeiffer in stilettos and clawed black gloves in the 1992 film *Batman Returns*, whose Catwoman costume spawned slinky offshoots by Gianni Versace.

Superheroes can also inspire real-world science. The Flash, created in 1940, possessed the power of super speed, as symbolized by his sleek scarlet bodysuit. Several outfits on display may increase the speed of the wearer. The outer texture of Speedo's Fastskin FS-Pro swimsuit mimics shark skin, which the company claims reduces drag by around 4%. More impressive is Dava Newman's body-hugging, flexible BioSuit, a space suit that relies on the mechanical counter-pressure provided by tight layers of material to protect the wearer from the vacuum of space. Newman, a professor of aeronautics and astronautics at the Massachusetts Institute of Technology, intends the BioSuit to replace bulkier, gas-pressurized space suits.

Most impressive are the wing suits developed by Atair Aerospace. A pilot strapped to the rigid wing suit — two polyethylene wings filled with jet fuel, powering turbines that provide almost 500 newtons of thrust — can fly at speeds of up to 350 kilometres per hour. Because the wing suit's wearer cannot be detected by radar, the company is now developing a military model with which spies could jump out of an aeroplane in one country and fly to another. Inventor Daniel Preston, the founder of Atair, says he has sky-dived hundreds of times in either the rigid wing suit or a non-fuelled soft suit, which has fabric webbing between the legs and arms. No other experience so exactly captures life as a superhero, Preston says. "It's as close as you can get to being a bird." ■

Josie Glausiusz is a writer based in New York.

## Super clothes with special powers

### Superheroes: Fashion and Fantasy

The Metropolitan Museum of Art,  
New York

Until 1 September 2008.

Shazam! With a bolt of lightning, 12-year-old Billy Batson turns into Captain Marvel, a superhero with the wisdom of Solomon, the strength of Hercules, the stamina of Atlas, the power of Zeus, the courage of Achilles and the speed of Mercury — legendary heroes whose initials spell the magic command that gives Marvel his superhuman powers. With a similar spell, the Metropolitan Museum of Art in New York has transformed one of its galleries into a shrine to modern mythical titans. Its new exhibition, *Superheroes: Fashion and Fantasy*, is craftily planted in the midst of its Greek and Roman art collection. Marble statues of Hercules, Diana and Perseus along with amphorae depicting muscular runners and wrestlers surround their fantastic descendants: Superman, Wonder Woman, Iron Man and The Incredible Hulk.

Extending our fascination with extremes of strength, endurance, speed and courage, the exhibition shows how the exaggerated forms of superheroes are mirrored in *haute couture*. It also demonstrates how inventors have incorporated aspects of superheroism — elasticity, rigidity and aerodynamic grace — into more practical kinds of clothing, such as



Sky-dive like a superhero in Atair's soft wing suit.

swimsuits, space suits and wing suits.

Superheroes might be mutants, armoured men, shape-shifters or gadgeteers; fashion designers draw inspiration from them all. Mutants — usually the result of a lab accident, genetic mishap or nuclear bomb blast — often appear in near-monstrous forms, such as The Incredible Hulk. Designers have transmuted these creations into garments of

# ESSAY



## Lost in music

Music provides unique opportunities for understanding both brain and culture. But globalization means that time is running out, warns **David Huron**, for the quest to encounter the range of possible musical minds.

**N**ever has so much music been so easily available. Go online, and you can download millions of recordings, from Spanish flamenco to Inuit throat singing. As a consequence, people are aware of the diversity of 'world musics' as never before.

But this rich cacophony is the soundtrack to a collapse in the diversity of musical minds. A Nigerian group might sing in Yoruba, but the harmonies are thoroughly Western. Native American Navajo singers make valiant efforts to preserve their traditions, but to the trained musicologist, their singing bears the unmistakable imprint of Western scales. The casual listener hears a wealth of variety; the musicologist detects a rapidly spreading monoculture — albeit expressed in many forms.

Music scholars have long been aware of the homogenizing effects of globalization<sup>1</sup>. Of course, musical cultures have always hybridized. The Silk Road, which connected Asia with the Mediterranean for nearly 2,000 years, had marked impacts on the music of Persia and Mesopotamia. The Atlantic slave trade brought people from West and Central Africa to the Caribbean and the Americas for 300 years, and the vibrant musical consequences of this human tragedy are all around us.

Today, one musical culture, that of the West, is influencing all others. What do we risk losing? Well, suppose that we find a musical behaviour present in all the world's cultures. This could reveal some universal in human behaviour. But if all the world's musics are influenced by a single dominant culture, universals become uninterpretable. A behaviour might be an innate cognitive disposition, or just an artefact of westernization. We won't be able to work out, for example, whether people

in different cultures perceive dissonance — an unpleasant combination of notes — in a similar way, or whether similar responses arise from exposure to Western music.

As the diversity of musical minds disappears, researchers will increasingly turn to Plan B: mining the recorded archives, assembled over the past century by the heroic efforts of ethnomusicologists. Fortunately, much of this was recorded before westernization took its toll. Unfortunately, Plan B looks less tenable than previously thought. The situation is alarming to those studying the cognitive neuroscience of music.

### Spandrel or foundation?

Music provides unique opportunities for understanding both brain and culture. Scientifically, we know relatively little about the peculiar human obsession with music. Perhaps music is a spandrel — an artefact of the developmental foundations of language. Or perhaps music has a unique phylogenetic origin<sup>2</sup>. We don't know. Emotionally, music and language seem to share a single code: a pitch contour that sounds sad when spoken will also sound sad played on an instrument<sup>3</sup>. Yet other research suggests that the neural mechanisms involved in rhythm are unrelated to language<sup>4</sup>.

There are innumerable pitfalls to understanding music and musical experience. Consider a simple aspect of melodic organization. Like the movements of the stock market, the up-and-down meandering of melodies has been the object of sustained statistical study<sup>5</sup>.

In many cultures, including Western, melodic pitches are normally distributed, and like all values drawn from such distributions, successive values regress towards the mean. When you encounter a tall person on the street, you might successfully predict that the next person you encounter will be shorter. But the presence of a tall person did not cause the next person to be shorter. The operative principle is simply that most people are of average height.

Something similar happens in melodies. But Western-enculturated listeners, anticipating whether the next pitch will be higher or lower, do not appreciate this. Instead, they expect large changes of pitch in a melody to be followed by a change of direction (this is called post-skip reversal). Listeners do not expect regression-to-the-mean even though this is the underlying principle<sup>6</sup>.

This, and similar research with Western listeners has taught us an important lesson: the objective organization of sounds is only loosely related to how minds interpret those sounds. A piece of music may exhibit features A, B and C, but only careful experimentation will show that listeners interpret A as

X, hear B imperfectly as B2, and are completely oblivious of feature C. Western melodies, for example, have an objective tendency to rise and then fall in pitch. Although enculturated listeners expect the ends of melodies to descend, they are curiously insensitive to the initial ascent. For centuries, Western music scholars wrongly assumed that common objective patterns in the scores were directly apprehended by listeners. If thoughtful West-

**"Even in the western Amazon, people listen to Funk Carioca and Christina Aguilera."**





**D. PARKINS** erners could be so wrong about interpreting their own music, imagine the capacity for self-deception regarding the music of another culture. When we observe an objective pattern in the music of some culture, we cannot assume that the pattern has any significance for culturally knowledgeable listeners.

Like linguists trying to make sense of sound recordings from an extinct language, music psychologists have realized that the archives of ethnomusicological field recordings will tell us little about the minds behind those musical cultures. One cannot do experimental studies with pre-existing data, and so causality cannot be inferred. Correlational studies are no substitute for true experimental manipulation.

### Difference engine

Variance is the lifeblood of empirical research. Without variability, data tell us little. In the case of musical behaviours, such variability has been found between the sexes, according to age and with respect to musical training. But these experiments have been largely limited to the lab rat of psychology — Western undergraduate students. Music psychologists have belatedly realized the importance of carrying out experiments with rapidly disappearing non-Western cultures. We don't yet know whether cultural differences dwarf the differences we see within Western culture.

When Paul Ekman did his classic studies of human facial expressions<sup>7</sup>, he rightly sought out people who had limited contact with Western people, movies and even photographs. Working with isolated cultures was essential, otherwise any behavioural similarities could be discounted as artefacts of cultural contamination. Comparable cross-cultural

experiments in sound are rare. In fact, few of the most basic musical concepts proposed by scholars have been tested in non-Western cultures.

Last year I joined an expedition of biologists to the remote Javari region of the Amazon. The biologists were censusing the wildlife. I was interested in the people. We encountered subsistence hunter–farmers with transistor radios. Even in the western Amazon, people listen to Funk Carioca and Christina Aguilera.

Linguists know how fast languages disappear. Musical cultures may be an order of magnitude more fragile. It will be many centuries before the whole world speaks Mandarin. Meanwhile Western music has swept the globe faster than aspirin. Robust musical cultures remain in China, India, Indonesia and the Arab world, but even in these regions, most people are thoroughly acquainted with Western music through film and television. Less robust musical cultures are disappearing rapidly or are showing deep infiltration by Western musical foundations. Many have already disappeared. There remain only a few isolated pockets, such as the highlands of Papua New Guinea and Irian Jaya.

Regrettably, most cognitive scientists are ill-equipped to do remote field work, and few ethnomusicologists know how to do an experiment. This situation must change rapidly if we are to have much hope of glimpsing the range of possible musical minds. We have perhaps just a decade or so before everyone on

the planet has been brought up with Western music or its derivatives.

Of course, we shouldn't underestimate future researchers' methodological cleverness in separating hybrid cultural experiences into their prior constituents. And it may be that all of the important lessons to learn about music can be found in Western music. But it would be rash to rely on these hopes.

In future centuries, music scholars may well curse our generation. We have the technical means to study different musical cultures and we still have a few isolated cultures to study. In the long span of music research, we live at a unique but fleeting moment. ■

**David Huron** is at the School of Music & Center for Cognitive Science, Ohio State University, Columbus, and author of *Sweet Anticipation: Music and the Psychology of Expectation*.

1. Nettl, B. *The Study of Ethnomusicology: Twenty Nine Issues and Concepts* (Univ. Illinois Press, Urbana, 1983).
2. Wallin, N. L., Merker, B. & Brown, S. *The Origins of Music* (MIT Press, Cambridge, Massachusetts, 1999).
3. Juslin, P. & Laukka, P. *Psychol. Bull.* **129**, 770–814 (2003).
4. Patel, A. D. *Music, Language, and the Brain* (Oxford Univ. Press, Oxford, 2007).
5. Temperley, D. *Music and Probability* (MIT Press, Cambridge, Massachusetts, 2007).
6. von Hippel, P. & Huron, D. *Music Percept.* **18**, 59–85 (2000).
7. Ekman, P. *The Face of Man: Expressions of Universal Emotions in a New Guinea Village* (Garland STPM Press, New York, 1980).

## LASER TECHNOLOGY

## Over the rainbow

John M. Lupton

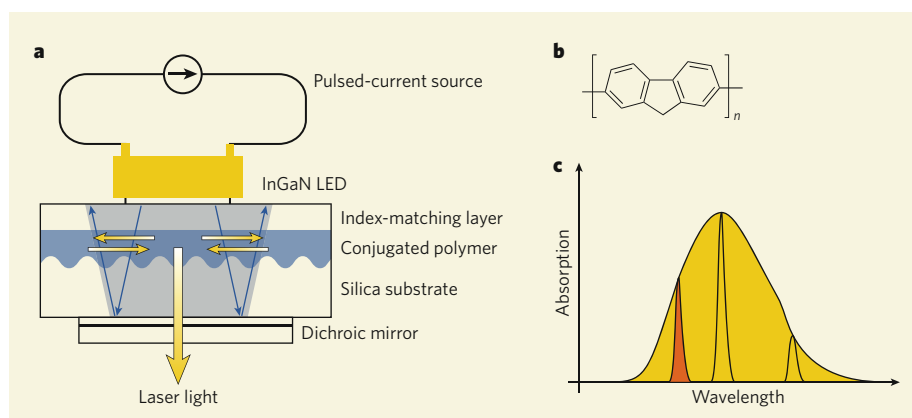
Many laser diodes provide light in only a limited range of the visible spectrum. A hybrid laser made out of plastic, driven by a high-power light-emitting diode, looks to offer a more flexible approach.

In the early days of semiconductor lasers, the choice of wavelengths was reminiscent of a famous Monty Python skit: it was a case of spam, spam or spam. The spectrum of available colours has since expanded impressively, but large gaps still exist, particularly at yellow wavelengths. Writing in *Applied Physics Letters*, Yang, Turnbull and Samuel<sup>1</sup> join up the dots, describing an ingenious laser that uses an inorganic light-emitting diode (LED) to activate an organic lasing material. This cheap and compact device promises an unbroken rainbow of lasing wavelengths for optical communications and analytical spectroscopy.

In its classic form, laser emission is brought about by pumping a medium with energy, either as light or as electric current. The aim is to heave — or ‘pump’ — so many atoms or molecules within the medium up from their ground state into an excited state that a population inversion is established, with more atoms in the higher-energy state than in the lower. Each excitation boosts an atomic electron into a higher energy level, leaving behind a positively charged hole where the electron used to be. Electron and hole recombine after a short while, and stimulate others to follow suit. The result is the emission of amplified, coherent light of a single wavelength.

Plastics have long seemed to hold promise as lasing materials<sup>2–4</sup>, largely because of their structure — or rather, their comparative lack of it. Inorganic semiconductors such as gallium arsenide, which are traditionally used as lasing media, have rigid atomic lattices with long-range order. Charge carriers can therefore wander through them relatively unimpeded, making pumping them using electric current easy. The downside is that the wavelengths of optical transitions in these materials are equally rigidly fixed. The rather disordered structure of plastic semiconductors, on the other hand, can be synthesized with widely varying optical and electronic properties.

During the past two decades, four main applications of plastic semiconductors have been identified: organic light-emitting diodes (OLEDs); solar cells; field-effect transistors (the bedrock of integrated circuits); and lasers. Of these, only lasers have so far resisted serious



**Figure 1 | Hybrid technology.** **a**, In Yang and colleagues' hybrid laser system<sup>1</sup>, a bright, pulsed inorganic (indium gallium nitride) light-emitting diode (LED) pumps light into an organic laser structure. A thin layer bridges the refractive-index gap between the LED and the organic semiconductor (a conjugated polymer) beneath, minimizing refractive losses. Light generated in the polymer bounces back and forth in the plane of the film owing to reflections from a periodically corrugated silica substrate. This provides optical feedback, and thus the gain necessary for laser action. A dichroic mirror reflects pump light back into the laser medium, while allowing laser light of the specific wavelength produced by the medium to leave. This wavelength can, in principle, be tuned continuously by varying the polymer material and the corrugation period of the grating<sup>9</sup>. **b**, The alternation of single and double carbon bonds in the paired phenylene rings of the lasing polymer allows electrons to move easily along the backbone, and thus produces efficient lasing. **c**, The disordered nature of the polymer plastic means that individual molecules have slightly different absorption spectra. Broadband LED light (yellow) can excite more molecules than narrow-band pumping by an external laser (orange), reducing the threshold for lasing, and promising cost and efficiency savings.

commercial exploration. Naively, one might assume that all one has to do to induce laser action in a plastic is to pump an OLED with sufficient power. So what's been holding us back?

Broadly speaking, three things. First, plastics have comparatively poor charge-transport characteristics, and so large numbers of charge carriers — and very high currents — are needed to generate a population inversion through electrical pumping<sup>3,4</sup>. To add insult to injury, the presence of this horde of charge carriers would impede the electron–hole recombination by which laser light is generated.

Second, laser action requires the use of mirrors at the boundaries of the laser medium to reflect light to and fro, and thus to build up sufficient intensity gain. Because OLEDs are extremely thin, the device's metal electrodes interfere with this 'optical feedback'.

But the third, and most daunting, obstacle to lasing OLEDs is that much of the energy

they generate is funnelled into particular electron–spin states known as triplet excitations<sup>5</sup>. Triplets are 'dark states', the nemesis of molecular photophysics. An electron can fall back into a hole and emit a photon only if the electron and hole spins match up; in triplet states, this isn't the case, and radiative recombination is forbidden. Long-lived triplets cause photobleaching — the chemical destruction of the surrounding emitting structure — and quench laser action in conventional lasers (if that weren't enough indication of villainy, they have also been implicated as a cause of skin cancer<sup>6</sup>). Triplets arise through strong quantum-mechanical interactions on the small length scales characteristic of OLED materials; in larger systems such as semiconductor crystals, their effects are negligible.

Hence the impetus behind Yang and colleagues' development of a hybrid device, the two separate components of which play to the



fortes of both inorganic semiconductors (ease of light generation) and organic semiconductors (flexibility in the wavelength generated). First, a high-power inorganic LED — unconventionally operated in a pulsed mode with its focusing lens removed — generates incoherent, spectrally broad light. That light is then converted into coherent radiation in an organic, plastic lasing medium situated immediately beneath the LED (Fig. 1a). For this medium, the authors chose a conjugated polymer derived from polyfluorene, with a backbone consisting of paired phenylene rings (Fig. 1b). The characteristic alternation of single and double covalent (shared-electron) bonds in this hydrocarbon chain means electrons can move along it efficiently, such that its response to the optical pumping from the LED is strong.

The new device is more compact and much cheaper than plastic lasers pumped with inorganic laser diodes<sup>7,8</sup>. Whereas such diodes emitting blue or ultraviolet light come with price tags of hundreds of dollars, high-power LEDs (which are also increasingly edging out traditional incandescent bulbs for lighting applications) are available for just cents. But that's not the best of it: because plastics are inherently disordered, made up of polymer chains jumbled up like a plate of spaghetti, different units on a chain emit light of slightly different colours. The absorption spectrum of the whole ensemble is made up of a superposition of narrower transitions corresponding to these units (Fig. 1c). Whereas a narrow-band pump laser will excite only a small subset of the molecules available, an LED with a broad emission spectrum can shovel more optically active units into the excited state, potentially lowering the threshold power needed to stimulate lasing.

By changing the laser medium and varying the corrugation of the silica substrate on which the device rests, it will be easy to tune such a laser system across the visible spectrum<sup>9</sup>. Plastics are not good conductors of heat, and so plastic lasers are unlikely to provide high power output, but in many applications — biomedical diagnostics and optical communications<sup>10</sup>, to name but two areas — precise wavelength control trumps brute force. The lasing future of plastics might not be as bright as that of other materials; but it certainly promises to be more colourful.

John M. Lupton is in the Department of Physics, University of Utah, Salt Lake City, Utah 84112, USA. e-mail: lupton@physics.utah.edu

## CELL BIOLOGY

## Two hands for degradation

Yasushi Saeki and Keiji Tanaka

**Living cells must do away with regulatory proteins that are not needed. News comes of a considerable advance in understanding how the main agent of destruction, the proteasome, catches its targets.**

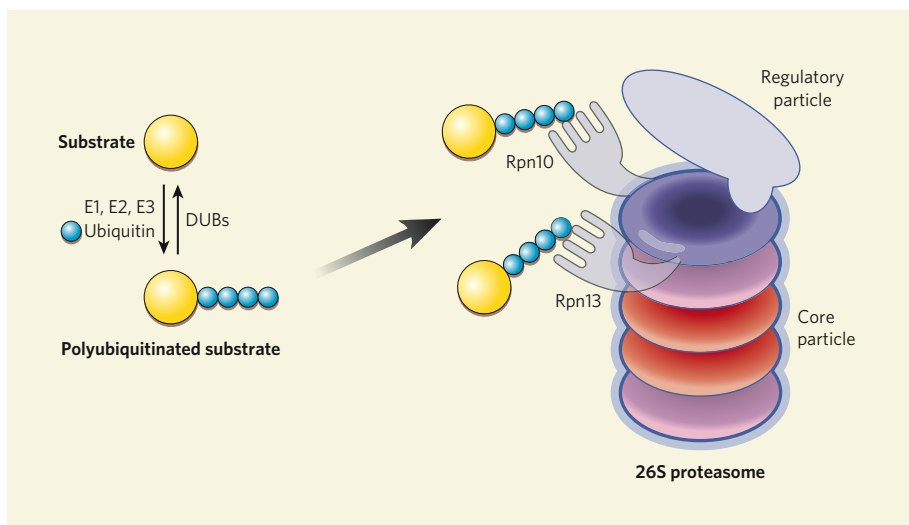
The 26S proteasome is a formidable piece of equipment — it is one of the principal cellular machines for carrying out the essential task of degrading proteins. Proteins to be destroyed are marked with tags in the form of the small protein ubiquitin, and when the proteasome encounters such polyubiquitinated proteins, it catches, then degrades them. Papers by Husnjak *et al.*<sup>1</sup> and Schreiner *et al.*<sup>2</sup>, which are the fruits of a multi-group collaboration and appear on pages 481 and 548 of this issue, show that the proteasome has, not one, but at least two hands with which it latches on to its ubiquitinated prey.

The ubiquitin–proteasome system controls almost all cellular processes — such as progression through the cell-division cycle and signal transduction — by degrading regulatory proteins<sup>3</sup>. The long journey to the destruction of a protein is started by covalent tagging with a chain consisting of several copies of ubiquitin, through the concerted action of a cascade of enzymes. Principally, polyubiquitin chains that consist of up to four or more ubiquitin molecules serve to promote degradation by the 26S proteasome. This protein is a multi-catalytic enzyme, with a highly ordered structure that is composed of at least 33 different subunits arranged in two sub-complexes — a 20S core particle and one or two 19S regulatory particles. The protein-degrading sites lie inside the

core particle and are accessible only through a narrow channel, so substrate proteins must be unfolded to reach the sites. The regulatory particle recognizes the polyubiquitin chains and removes them, then unfolds the substrate proteins and transfers them into the core particle for destruction.

How the polyubiquitinated proteins are recognized by the proteasome is a fundamental and long-standing question. In 1994, Rpn10, one subunit of the regulatory particles, was identified as a protein that binds to polyubiquitin chains; it does so via a ubiquitin-interacting motif (UIM) found at one end of the protein (the carboxy terminus)<sup>4,5</sup>. Genetic experiments in yeast, however, showed that deletion of the *RPN10* gene or a *uim* mutation had few or no effects. These results raised the possibility that other proteasomal ubiquitin receptors exist that can compensate for Rpn10 function.

Several laboratories pursued this possibility and identified proteins with particular structural units — ubiquitin-like proteasome-binding domains (UBL) and ubiquitin-associated domains (UBA) — as being implicated in targeting ubiquitin. The proteins concerned included Rad23, Dsk2 and Ddi1 (refs 5, 6). The finding that *RAD23* and *DSK2* interact genetically with the *rpn10* mutation, together with a subsequent biochemical study, established that the UBL–UBA-containing proteins function



**Figure 1 | A pair of hands for catching ubiquitin.** Protein substrates are marked for degradation by polyubiquitination, which is carried out by E1 (activating), E2 (conjugating) and E3 (ligating) enzymes; deubiquitinating enzymes (DUBs) can reverse this process. If it is not reversed, the ubiquitin units are recognized by the 26S proteasome protein-degrading machine through two intrinsic receptors, Rpn10 and the newly identified<sup>1,2</sup> Rpn13. Extrinsic ubiquitin receptors, such as Rad23, Dsk2 and Ddi1 (not shown), also function cooperatively in this process.

1. Yang, Y., Turnbull, G. A. & Samuel, I. D. W. *Appl. Phys. Lett.* **92**, 163306 (2008).
2. Tessler, N., Denton, G. J. & Friend, R. H. *Nature* **382**, 695–697 (1996).
3. McGehee, M. D. & Heeger, A. J. *Adv. Mater.* **12**, 1655–1668 (2000).
4. Kozlov, V. G. *et al. IEEE J. Quantum Electron.* **36**, 18–26 (2000).
5. Reufer, M. *et al. Nature Mater.* **4**, 340–346 (2005).
6. Hiraku, Y. *et al. Photochem. Photobiol.* **83**, 205–212 (2007).
7. Riedl, T. *et al. Appl. Phys. Lett.* **88**, 241116 (2006).
8. Karnutsch, C. *et al. IEEE Photon. Technol. Lett.* **19**, 741–743 (2007).
9. Riechel, S. *et al. Opt. Lett.* **26**, 593–595 (2001).
10. Xia, R., Heliotis, G. & Bradley, D. D. C. *Appl. Phys. Lett.* **82**, 3599 (2003).

as extrinsic ubiquitin receptors of the proteasome<sup>5,7</sup>. Thus, the question of ubiquitin receptors seemed to be answered. As we now find out, however, the 26S proteasome concealed an additional intrinsic ubiquitin receptor.

In the first of the new papers, Husnjak *et al.*<sup>1</sup> describe how they have identified human Rpn13, a regulatory-particle subunit, as a ubiquitin-binding protein. Although both the amino- and carboxy-terminal regions of Rpn13 are conserved among species, the ubiquitin-binding activity is located at what is known as a pleckstrin-homology-like domain at the amino terminus (pleckstrin-homology domains are common in proteins involved in intracellular signalling). Rpn13 from budding yeast has only the amino-terminal conserved domain.

Husnjak *et al.*<sup>1</sup> first addressed the significance of the ubiquitin-binding activity of Rpn13 in purified 26S proteasomes. Although proteasomes lacking all known ubiquitin-receptor activities — including the UIM of Rpn10 and three UBL-UBA-containing proteins — still bound to the polyubiquitinated substrate, additional deletion of Rpn13 resulted in almost total loss of ubiquitin-binding activity. The defect was restored by either Rpn10 or Rpn13. These results clearly suggest that Rpn10 and Rpn13 are the primary ubiquitin receptors of the 26S proteasome (Fig. 1).

The amino-terminal domain of Rpn13 shows no similarity to known ubiquitin-binding motifs. As Husnjak *et al.*<sup>1</sup> and Schreiner *et al.*<sup>2</sup> recount, the next phase of the research was to use nuclear magnetic resonance and crystallographic studies to determine how Rpn13 binds ubiquitin. These structural analyses revealed that the amino-terminal domain has a canonical pleckstrin-homology fold consisting, in technical terms, of a seven-stranded  $\beta$ -sandwich structure capped by the carboxy-terminal  $\alpha$ -helix. The authors therefore named this domain 'pleckstrin-like receptor for ubiquitin' (Pru).

They found that the Pru domain of human Rpn13 shows high affinity (around 90 nanomolar) for diubiquitin, the strongest binding among the known ubiquitin receptors. Both human and yeast Rpn13 Pru domains use three loops at one edge of their  $\beta$ -sheet to bind ubiquitin. The authors successfully created an *rpn13* mutant (called *rpn13-KKD*) that lost ubiquitin-binding capacity without compromising proteasome integrity, and tested the biological effects of this mutation in yeast. Degradation of a model substrate protein of the ubiquitin-proteasome system was retarded in this mutant; and when combined with an *rpn10-uim* mutant, the cells showed further impairment of proteasome function. In addition, polyubiquitinated proteins accumulated in the *rpn10-uim*, *rpn13-KKD* mutant cells. These results suggest that Rpn13 is a true intrinsic ubiquitin receptor of the 26S proteasome, and that it collaborates with Rpn10 *in vivo*.

An obvious question that arises is why there are so many ubiquitin receptors in

the ubiquitin-proteasome system. The 26S proteasome binds with high affinity to the longer polyubiquitin chains, so it is likely that both Rpn13 and Rpn10 can bind simultaneously to a substrate that bears such chains. Rpn13 Pru can also recognize UBL-UBA-containing proteins<sup>1,2</sup>, as mammalian Rpn10 does<sup>4</sup>. Perhaps polyubiquitin recognition at multiple sites in the proteasome enhances targeting potency and stabilizes the proteasome-substrate complex for substrate degradation. Intriguingly, yeast cells with mutations in five ubiquitin receptors are still viable, indicating that there may still be unidentified ubiquitin receptors in the proteasome, perhaps operating downstream from Rpn10 and Rpn13. In mammalian cells, Rpn13 binds via its carboxy-terminal domain to Uch37, one of three proteasome-associated deubiquitinating enzymes<sup>8–10</sup>. This means that Rpn13 might be a specialized ubiquitin receptor that can fine-tune the timing of substrate degradation.

More generally, it is becoming apparent that there are several layers to proteasome regula-

tion, and that this may allow the proteasome to cope with high substrate flux as well as a wide diversity of substrates. The identification of Rpn13 as a ubiquitin receptor will help in directing research to elucidate these intricate mechanisms. ■

Yasushi Saeki and Keiji Tanaka are at the Tokyo Metropolitan Institute of Medical Science, 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan.  
e-mails: saeki@rinshoken.or.jp;  
tanakak@rinshoken.or.jp

1. Husnjak, K. *et al.* *Nature* **453**, 481–488 (2008).
2. Schreiner, P. *et al.* *Nature* **453**, 548–552 (2008).
3. Pickart, C. M. & Cohen, R. E. *Nature Rev. Mol. Cell Biol.* **5**, 177–187 (2004).
4. Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. *J. Biol. Chem.* **269**, 7059–7061 (1994).
5. Elsasser, S. & Finley, D. *Nature Cell Biol.* **7**, 742–749 (2005).
6. Wilkinson, C. R. *et al.* *Nature Cell Biol.* **3**, 939–943 (2001).
7. Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. *Cell* **118**, 99–110 (2004).
8. Yao, T. *et al.* *Nature Cell Biol.* **8**, 994–1002 (2006).
9. Hamazaki, J. *et al.* *EMBO J.* **25**, 4524–4536 (2006).
10. Qiu, X. B. *et al.* *EMBO J.* **25**, 5742–5753 (2006).

## BIOPHYSICS

# Cells get in shape for a crawl

Jason M. Haugh

**A cell's shape changes as it moves along a surface. The forward-thinking cytoskeletal elements are all for progress, but the conservative cell membrane keeps them under control by physically opposing their movement.**

The ability of living cells to move affects the way our bodies develop, fight off infections and heal wounds. Moreover, cell migration is an extremely complex process, which explains why it has captured the collective imaginations of a variety of fields, from the biological and the physical sciences. This is good news, because cell motility is determined in equal parts by biochemistry and mechanics<sup>1,2</sup>, and so understanding and manipulating it require the sort of clever approach that comes only from the integration of multiple scientific disciplines. On page 475 of this issue, Keren *et al.*<sup>3</sup> combine approaches familiar to cell biology with those familiar to applied mathematics and physics to address how the forces generated by specific molecular processes in a cell produce its observed shape.

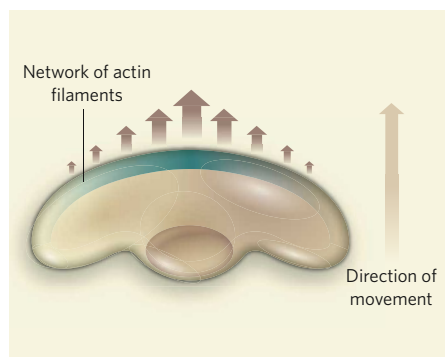
The starting point for the authors' analysis was the characterization of variability in the shapes adopted by epithelial keratocytes from fish skin in culture. These cells serve as a unique model system for studying cell migration, because they crawl rapidly and without frequent changes in direction, and maintain a nearly constant shape as they move. Their stereotypical shape, often described as an 'inverted canoe', is characterized by a broad membrane structure at its front, the lamellipodium, which

protrudes forward in concert with forces that act at the rear of the cell. The authors determined that most of the shape variability could be attributed to differences in cell size and, to a lesser extent, the aspect ratio of its characteristic dimensions (the ratio of its width to its height).

The key insight by Keren *et al.* was to relate two independent observations: the cell's shape and its distribution of actin filaments. Actin filaments are structural elements inside the cell that, through the energy-intensive process of adding (and later removing) protein subunits, produce the mechanical work required to push the cell forward. New, growing filaments are formed by the branching off of existing ones, a process that is well understood in keratocytes<sup>4,5</sup>.

Building on previous work<sup>6</sup>, the authors propose a mathematical model to explain the observation that the filament density at the cell front is graded, with the highest density at its centre (Fig. 1). The importance of this approach is that it incorporates known molecular mechanisms, and hence the model could be used to predict what might happen if the functions of the molecules involved were perturbed. The authors next invoked what is known as the force-velocity relationship, which states that the rate at which the





**Figure 1 | Shape matters.** Viewed from above, the characteristic shape of fish keratocyte cells crawling on a surface resembles an inverted canoe. The driving force of the cell's movement comes from actin filaments that form a network at the cell front. The filaments grow in the direction of motion, generating a thrust that overcomes tension in the cell membrane. Keren *et al.*<sup>3</sup> show that the density of actin filaments varies across the cell front (higher-density regions are shown in deeper turquoise). The authors propose that high-density regions generate more thrust than low-density regions (arrow sizes indicate magnitude of thrust). High-density regions thus protrude forward more than low-density areas. This model explains the shapes formed by moving cells.

membrane can be pushed forward by the growing actin filaments decreases as the force resisting them increases, and above a critical value — the stall force — protrusion stops completely.

Although the mechanisms that give rise to this relationship are actively debated, it is strongly grounded by empirical observations<sup>7</sup>. Keren *et al.*<sup>3</sup> reasoned that the load force per actin filament must increase as the filament density decreases from the centre of the cell, and thus the 'sides' of the cell represent the regions of the lamellipodium where the actin filaments are stalled (and/or buckled under pressure; Fig. 1). A specific prediction followed, which the authors confirmed: the steepness of the actin-filament gradient from the cell centre to the front edges is directly related to the cell's aspect ratio. Furthermore, with the specification of the cell shape and the force-velocity relationship, Keren *et al.* showed that they could predict, in a consistent way, the curvature of the cell front and the cell-migration speed.

The elegance of the authors' model, which exemplifies the combined use of quantitative cell biology and mathematical analysis<sup>8</sup>, lies in its ability to relate molecular and physical processes with very few or in some cases no adjustable parameters. One unresolved issue that warrants further study concerns the mechanistic implications for the variability in cell size. Although Keren *et al.* were not able to address this point directly, their model suggests that it ought to affect either the rate of actin-filament branching or the tension of the cell membrane, or possibly both. ■

Jason M. Haugh is in the Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695-7905, USA.  
e-mail: jason\_haugh@ncsu.edu

1. Lauffenburger, D. A. & Horwitz, A. F. *Cell* **84**, 359–369 (1996).

2. Ridley, A. J. *et al.* *Science* **302**, 1704–1709 (2003).
3. Keren, K. *et al.* *Nature* **453**, 475–480 (2008).
4. Pollard, T. D., Blanchoin, L. & Mullins, R. D. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576 (2000).
5. Pollard, T. D. & Borisy, G. G. *Cell* **112**, 453–465 (2003).
6. Lacayo, C. I. *et al.* *PLoS Biol.* **5**, e233 (2007).
7. Mogilner, A. *Curr. Opin. Cell Biol.* **18**, 32–39 (2006).
8. Mogilner, A., Wollman, R. & Marshall, W. F. *Dev. Cell* **11**, 279–287 (2006).

## ASTRONOMY

# Supernova bursts onto the scene

Roger Chevalier

**The stellar explosions known as supernovae are spectacular but common cosmic events. A satellite telescope's chance observation of a burst of X-ray light might be the first record of a supernova's earliest minutes.**

Once the processes of nuclear fusion that have bolstered it against its own gravity are exhausted, the core of a massive star collapses in on itself. The result is a cataclysmic explosion that sends a violent shock wave racing outwards. As this shock wave reaches the star's surface, it produces a short, sharp burst of X-ray or ultraviolet radiation, the prelude to the expulsion of most of the star's matter into the surrounding medium. Lasting days to months, we see this aftermath of the explosion as a supernova.

That is the theory, at any rate. But although supernovae themselves are common enough, the chain of events that lead up to them — in particular, the exact moment of 'shock break-out' — had never been seen. That all changes with a report from Soderberg *et al.* (page 469)<sup>1</sup>. They observed an intense, but short-lived, X-ray outburst from the same point in the sky where shortly afterwards a supernova flared up, and have thus provided valuable support for the prevalent theories of supernova progenitors.

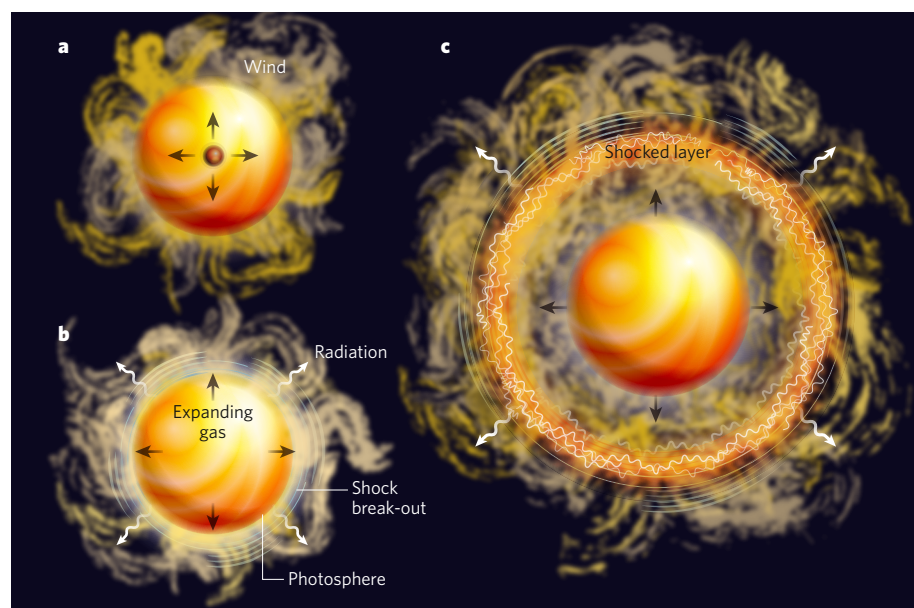
The authors' discovery was serendipitous: they just happened to be examining the aftermath of a similar supernova, of 'type Ibc', in the same galaxy. The instrument they were using, NASA's Swift satellite, was primarily intended to pinpoint the mysterious flashes of intense, high-energy light known as γ-ray bursts. But, while pursuing this successful main career, the telescope has also developed a useful sideline in X-ray and optical follow-up observations of supernovae.

What Swift spotted<sup>1</sup> was an X-ray outburst that lasted for some 10 minutes. Its energy content was around  $10^{39}$  joules, about a hundred-thousandth of the energy expelled in the explosive motions of a supernova. Continued observation of the position of the outburst showed the emergence of a spectrum and an evolution of emission intensity over time typical of a type-Ibc supernova, albeit with a slightly fainter peak luminosity than normal.

The exploding object was also detected by NASA's Chandra X-ray observatory 10 days after the X-ray outburst, as well as in a series of radio measurements between 3 and 70 days after. Similar observations characterize type-Ibc supernovae, and are thought to relate to interaction of the expanding supernova with mass lost from its progenitor before the explosion, which encircles the star as a surrounding 'wind' (Fig. 1). The interaction generates shock waves that accelerate electrons to almost light speed. These electrons in turn emit radio-frequency synchrotron radiation as their paths curve in the ambient magnetic field, and scatter photons from the visible surface of the star, the photosphere, up to X-ray energies.

Taken together, these observations seem to add up to the identification of the X-ray outburst with the supernova — now designated SN 2008D — that followed. One caveat is that, although the energy of the outburst was close to predictions for the shock break-out of a type-Ibc supernova<sup>2</sup>, its duration was much longer than expected. The length of the burst should be determined by the time light needs to cross the supernova progenitor, which is 10 seconds or less. The implication, therefore, is that the photosphere of the progenitor star extends farther than expected, perhaps because it has shed a large amount of material before the supernova occurs.

Within the star, the energy behind the shock wave emanating from the core's collapse is dominated by radiation. Outside, it is dominated by gas energy. Shock break-out occurs at the transition between these two modes, when the radiation behind the internal shock wave spreads out into the circumstellar medium and accelerates its gas. As the inner, already accelerated layers of gas catch up with outer, slower-moving layers, an external gas shock wave develops. Soderberg *et al.*<sup>1</sup> suggest that the observed spectrum of the X-ray burst is determined by the shock acceleration of photons from the supernova photosphere. Detailed



**Figure 1 | Shock break-out.** Soderberg and colleagues' observations<sup>1</sup> of an X-ray outburst preceding a type-Ibc supernova provide sterling support for prevalent models of how these cosmic explosions occur. **a**, Once its nuclear fuel is exhausted, the core of a massive star collapses in on itself, generating a huge explosion that propagates outwards as a shock wave. The progenitor star is surrounded by a 'wind' of gas previously lost from the star. **b**, After a matter of minutes, the shock wave reaches the surface of the star (the photosphere) and radiation from the explosion that is trailing in its wake escapes, accelerating the surrounding gas outwards. The moment of this 'shock break-out' is what the authors succeeded in capturing. **c**, Days later, a layer of hot gas has developed where the rapidly expanding, but cool, supernova gas impacts on the surrounding wind. This shocked layer is the site of electrons moving close to the speed of light that are responsible for radio and X-ray emission, typical supernova signatures.

hydrodynamic simulations that allow for the effects of a radiation field that is out of equilibrium will be needed before we can properly model the niceties of the transition.

Is there any other possible interpretation of the X-ray outburst, other than the emergence of a supernova shock? Might the outburst simply be a lower-energy cousin of a  $\gamma$ -ray burst? These bursts are thought to be produced by the same type of progenitor as type-Ibc supernovae, albeit less than 100 times as frequently. Their pathology is very different: rather than being the result of a spherical shock wave rippling through the star, they are assumed to be caused by directed, relativistic jets of particles and magnetic fields that are generated by a central black hole or neutron star and then burrow through their surrounds.

The X-ray outburst from SN 2008D was much weaker than a  $\gamma$ -ray burst, although that might simply represent a downwards extension of the permissible intensity range<sup>3,4</sup>. It has also recently been suggested<sup>4</sup> that the outburst shares with  $\gamma$ -ray bursts certain relationships between the amount of energy radiated isotropically (that is, equally in all directions), its peak spectral energy and the peak luminosity of the ensuing supernova. But speaking against the  $\gamma$ -ray-burst interpretation is not only the weight of post-burst observations seeming to indicate a normal supernova, but also the lack of firm evidence for the relativistic motions that are a signature of  $\gamma$ -ray bursts.

The observations<sup>1</sup> of SN 2008D would thus seem to be the earliest of light emanating from a supernova, just minutes after core collapse. NASA's Galaxy Evolution Explorer (GALEX) has recently seen the rising ultraviolet emission associated with shock break-out from type-II supernovae<sup>5,6</sup>. The progenitors of these supernovae are surrounded by an envelope of hydrogen gas, and the shock wave takes close to a day to traverse them.

As well as telling us more about the types of star that produce supernovae, such observations of shock break-out, by helping to tie down the time of core collapse, could provide useful auxiliary information in terrestrial hunts for exotic citizens of the cosmos thought to be produced in these cosmic cataclysms. The benefits could be felt by neutrino detectors, and by detectors searching for evidence of the elusive ripples in space-time known as gravitational waves. ■

Roger Chevalier is in the Department of Astronomy, University of Virginia,

PO Box 400325, Charlottesville, Virginia 22904, USA.

e-mail: rac5x@virginia.edu

1. Soderberg, A. M. *et al.* *Nature* **453**, 469–474 (2008).
2. Matzner, C. D. & McKee, C. F. *Astrophys. J.* **510**, 379–403 (1999).
3. Xu, D., Zou, Y.-C. & Fan, Y.-Z. preprint at <http://arxiv.org/abs/0801.4325> (2008).
4. Li, L.-X. preprint at <http://arxiv.org/abs/0803.0079> (2008).
5. Schawinski, K. *et al.* preprint at <http://arxiv.org/abs/0803.3596> (2008).
6. Gezari, S. *et al.* preprint at <http://arxiv.org/abs/0804.1123> (2008).

## STRUCTURAL BIOLOGY

# Snapshots of DNA repair

Stephen C. Kowalczykowski

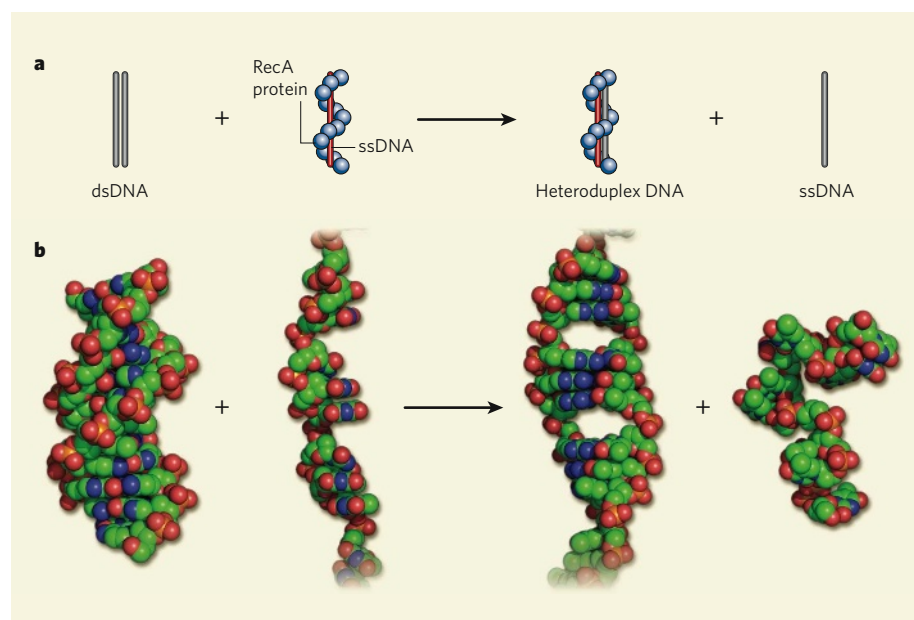
**In recombinational DNA repair, nearly identical sequences in chromosomes are found and swapped. Structures of the RecA–DNA complexes involved provide insight into the mechanism and energetics of this universal process.**

Homologous recombination is one of the many processes used by cells to repair damaged DNA and to diversify their genomes. A central step in recombination involves the exchange of DNA strands between identical, or nearly identical, segments of chromosomes. This crucial reaction is catalysed by the RecA family of DNA-strand-exchange proteins, which include the founding member in bacteria, Rad51 in eukaryotes and RadA in archaeans. On page 489 of this issue, Chen *et al.*<sup>1</sup> describe structures of both the substrate (RecA complexed with single-stranded DNA) and the product (RecA complexed with double-stranded DNA) of DNA strand exchange. The structures reveal non-uniform DNA stretching, and suggest a mechanism for strand exchange.

On the face of it, DNA strand exchange is a simple reaction: one strand of double-stranded DNA (dsDNA) is replaced with an identical single strand of DNA (Fig. 1a). This reaction would be simple, were it not for the stability of dsDNA, which resists strand separation, and the need to accurately align DNA sequences. Thus, although the stability is useful for the storage of genetic information, it is an impediment to DNA repair and recombination.

The RecA-like proteins do not deal with this problem by unwinding dsDNA, as helicase enzymes do. Instead, they first assemble on a single strand of DNA, which was generated in the preceding step of recombination, to form a helical nucleoprotein (protein–DNA) filament. When formed with ATP, this filament (termed the presynaptic complex) is the active





**Figure 1 | DNA strand exchange promoted by RecA protein.** **a**, The prototypical DNA-strand-exchange reaction. Double-stranded DNA (dsDNA) pairs with the RecA presynaptic filament, which consists of RecA protein and single-stranded DNA (ssDNA), to produce heteroduplex DNA bound by RecA and the exchanged ssDNA. **b**, Structures of the participating DNA molecules: B-form dsDNA; ssDNA within the presynaptic filament (as determined by Chen and colleagues<sup>1</sup>, protein not shown); dsDNA within the filament<sup>1</sup> (protein not shown); and randomly coiled ssDNA.

species in DNA strand exchange that searches for a homologous sequence within the dsDNA. Once found, DNA strand exchange occurs as a concerted swap of DNA strands. The hydrolysis of ATP inactivates the filament, and permits disassembly of the complexes<sup>2</sup>.

Herein lie the mysteries of DNA strand exchange. How does the RecA nucleoprotein filament recognize DNA sequence identity? And, on finding it, how does the exchange occur? How is the stability of dsDNA overcome? Partial answers to these questions emerged from biochemical studies. The homology search is a 'simple' collisional process because ATP hydrolysis is not essential, only ATP binding. In fact, ATP binding by the RecA nucleoprotein filament is sufficient for DNA strand exchange<sup>3</sup>. The free energy of ATP hydrolysis is not directly involved in the exchange of DNA strands; rather, the free energy of presynaptic-filament binding to the dsDNA 'activates' it by extending and untwisting it, making the duplex DNA a willing participant in the exchange process<sup>4</sup>. ATP hydrolysis then allows dissociation of all participants: a classic case of 'credit card' energetics (expend now, pay later ...).

Structural information derived from electron microscopy was particularly revealing with respect to these questions. The ATP-bound form of the RecA nucleoprotein filament is extended by about 50% relative to standard B-form DNA, with around 6.2 RecA monomers and 18 DNA base pairs per turn<sup>5</sup>; this extended filament is also seen for all RecA homologues<sup>6</sup>. In contrast, the DNA in the inactive ADP-bound nucleoprotein filament is less extended. Thus, the RecA nucleoprotein

filament undergoes ligand-induced structural transitions between an active, extended filament and an inactive, compact filament<sup>5</sup>. The electron microscopy studies also revealed that the RecA nucleoprotein filaments are structurally polymorphic, varying in pitch, width and extension<sup>5</sup>, highlighting the challenge facing higher-resolution structural analysis.

Nonetheless, in 1992 the crystal structure of a RecA filament was solved<sup>7</sup>, but the structure lacked DNA and was of the inactive ADP-bound compact form<sup>8</sup>. For many years, the absence of a structure of the active form confounded molecular and mechanistic interpretation<sup>9</sup>.

However, Chen and colleagues<sup>1</sup> now elucidate structures of RecA assembled on single-stranded DNA (ssDNA) and on dsDNA, defining both the substrate and product forms of the reaction, respectively. How did they succeed? The authors recognized that the intrinsic conformational flexibility of the RecA nucleoprotein filaments and their capacity to self-assemble indefinitely might hinder crystallization in the active state. Their solution to these problems was ingenious, and is applicable to other self-assembling systems.

First, they created 'pre-polymerized' assemblies of RecA protein by fusing four, five or six monomers of RecA into a single polypeptide chain. To prevent indefinite polymerization of the resulting 'mini-filaments', the sites for monomer-monomer interactions were deleted from the first and last monomers in the chain. Despite the many potential pitfalls, this approach worked splendidly, producing functional proteins. When assembled in the presence of an ATP analogue on DNA that exactly accommodated these fusion proteins (15 and



## 50 YEARS AGO

The launching of *Sputnik 3* (Satellite 1958  $\delta$ ) was announced from Moscow on May 15. The satellite was stated to be conical in shape, with a length of 12.3 ft. excluding aerials, a base diameter of 5.7 ft. and a weight of 2,926 lb., including 2,134 lb. of apparatus. The experiments for which the satellite is designed include studies of cosmic rays, geomagnetism, solar radiation and micrometeorites, and the results are to be telemetered back to the Earth. The satellite is equipped with solar batteries and carries a radio transmitter with a frequency of 20.005 Mc./sec. There are two other objects in orbit with the satellite, namely, the rocket which performed the last stage of propulsion and a nose cone which protected the instruments during the climb through the atmosphere.

From *Nature* 24 May 1958.

## 100 YEARS AGO

The last half-yearly number of the *Journal of the Royal Anthropological Institute* contains an important memoir, prepared by two enthusiastic Scotch anthropologists, Messrs. Gray and Tocher, on the pigmentation of hair and eyes among the school children of Scotland ... The highest density of fair hair is to be found in the great river valleys opening on the German Ocean and in the Western Isles. In the former case, this probably points to invasions of a blonde race into those regions. Similarly, the higher percentage of fair hair in the Spey valley and in the Western Isles implies inroads of the Vikings or Norsemen. It is perhaps pushing the evidence too far when the writers suggest that the high percentage of fair-haired girls in the neighbourhood of Dunfermline is due to the train of blonde damsels who are supposed to have accompanied the Saxon princess Margaret, who about the time of the Norman Conquest became Queen of Malcolm Canmore.

From *Nature* 21 May 1908.

50 & 100 YEARS AGO

18 nucleotides, respectively), these mini-filaments formed ordered crystals.

The structures reveal an ordered filament with 6.2 monomeric units per turn and a pitch of 92–95 Å. The DNA is close to the filament axis, is extended relative to B-form DNA, and has global features compatible with the electron microscopy. The ATP is completely buried at an interface between monomers. Each RecA monomer interacts with three nucleotides of the DNA (a triplet) adjacent to itself in the structure, as well as with two more nucleotides, one from each of the preceding and following triplets. As a result, each nucleotide triplet is bound by three monomers.

Perhaps the most remarkable feature of the nucleoprotein filament is the DNA structure (Fig. 1b). The 50% extension is not manifest as an isotropic extension at the nucleotide level; instead, the DNA is seen to comprise a three-nucleotide segment with a nearly normal B-form distance between bases (an axial rise of 3.5–4.2 Å for ssDNA and 3.2–3.5 Å for dsDNA), followed by a long untwisted internucleotide stretch (approximately 7.1–7.8 Å in ssDNA and 8.4 Å in dsDNA) before the next three-nucleotide element, and so on. This was a surprising result, because most people assumed that the DNA within the RecA–DNA complexes was uniformly stretched to an average of about 5.2 Å between bases.

The unusual repeat pattern of DNA extension in the RecA nucleoprotein filament offers a structural basis for understanding the dynamics of filament assembly. Assembly occurs by rate-limiting initiation of polymer formation (nucleation) followed by growth<sup>2</sup>. The structure shows that it would be energetically unfavourable for a single monomer to make the full repertoire of molecular contacts with DNA because of the need to both unstack the bases and extend the DNA to the next nucleotide triplet. Thus, the free energy for binding of the first monomer will be unfavourable relative to the binding of a second protein to an existing monomer, explaining the observed cooperativity of RecA binding to DNA<sup>2</sup>. Binding of a third monomer provides additional net free energy, because now two of the three monomers benefit from the added free energy of cooperative interactions. As more monomers bind, the energetic cost of extending the DNA is 'amortized' over an increasingly greater number of RecA monomers, until the net free energy of nucleus formation is sufficiently negative to permit stable nucleation. Although more complex scenarios can be envisaged, the structures of Chen and colleagues<sup>1</sup> now permit detailed energetic modelling of filament formation.

The results also highlight the physical mismatch between the ssDNA within the filament and the naked duplex DNA target (Fig. 1b). How does RecA align these sequences? The structures<sup>1</sup> offer provocative insights into how the transient three-stranded intermediate might look, and how the fidelity of DNA

strand exchange might be enforced. It is easy to imagine the pairing between an ssDNA triplet within the filament and the naked dsDNA, as both have approximately B-form dimensions. However, pairing of the next three base pairs of DNA requires extension of the dsDNA to conform to the observed extension of the ssDNA in the filament. This energetically unfavourable base unstacking and chain extension could be compensated both by the now lower entropic cost (because the next triplet is part of the already paired dsDNA) and by the favourable base-pairing interactions that would form if the next triplet were fully homologous. However, if even one of the base pairs was non-complementary, then the nascent paired molecule might not be sufficiently stable, and homologous pairing with a partially homologous sequence would be aborted.

Furthermore, the structures show that the strand complementary to the ssDNA in the presynaptic filament makes few contacts with the protein. Hence, it is largely stabilized by correct Watson–Crick base-pairing, thereby requiring accurate DNA pairing. Successful DNA pairing requires at least 15 base pairs of homology<sup>10</sup>, and the structures suggest how such fidelity is enforced.

Determination of the three-dimensional structure of the active state of RecA nucleoprotein filaments by Chen and colleagues<sup>1</sup> is a watershed in recombination biochemistry and mechanics. Not only do the structures inform us about this central protein, but they also

enable the formulation of structural hypotheses that relate to the RecA orthologues and to interacting proteins. Although the eukaryotic and archaeal RecA homologues differ in many functional and mechanistic details, the RecA structures will provide a valuable foundation for understanding them. Also, many proteins interact with the various forms of RecA family members to regulate assembly and disassembly of the filament. Having structures of both the ATP- and ADP–RecA nucleoprotein filaments will help clarify the mechanistic basis of their biological functions. Clearly, more (DNA) partner-swapping experiments will be forthcoming. ■

Stephen C. Kowalczykowski is in the Departments of Microbiology, and of Molecular and Cellular Biology, University of California, Davis, One Shields Avenue, Davis, California 95616, USA. e-mail: sckowalczykowski@ucdavis.edu

1. Chen, Z., Yang, H. & Pavletich, N. P. *Nature* **453**, 489–494 (2008).
2. Kowalczykowski, S. C. *Annu. Rev. Biophys. Chem.* **20**, 539–575 (1991).
3. Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. *Proc. Natl Acad. Sci. USA* **87**, 21–25 (1990).
4. Mazin, A. V. & Kowalczykowski, S. C. *Genes Dev.* **13**, 2005–2016 (1999).
5. Yu, X. et al. *Curr. Protein Pept. Sci.* **5**, 73–79 (2004).
6. Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. *Front. Biosci.* **3**, d570–d603 (1998).
7. Story, R. M., Weber, I. T. & Steitz, T. A. *Nature* **355**, 318–325 (1992).
8. Story, R. M. & Steitz, T. A. *Nature* **355**, 374–376 (1992).
9. McGrew, D. A. & Knight, K. L. *Crit. Rev. Biochem. Mol. Biol.* **38**, 385–432 (2003).
10. Ferrin, L. J. & Camerini-Otero, R. D. *Science* **254**, 1494–1497 (1991).

## CELL BIOLOGY

# Viruses in camouflage

Kirsten Sandvig and Bo van Deurs

**The vaccinia virus acts like a Trojan Horse to enter its host cells: it envelops itself in the membrane of a dying cell, and is then taken up by healthy cells.**

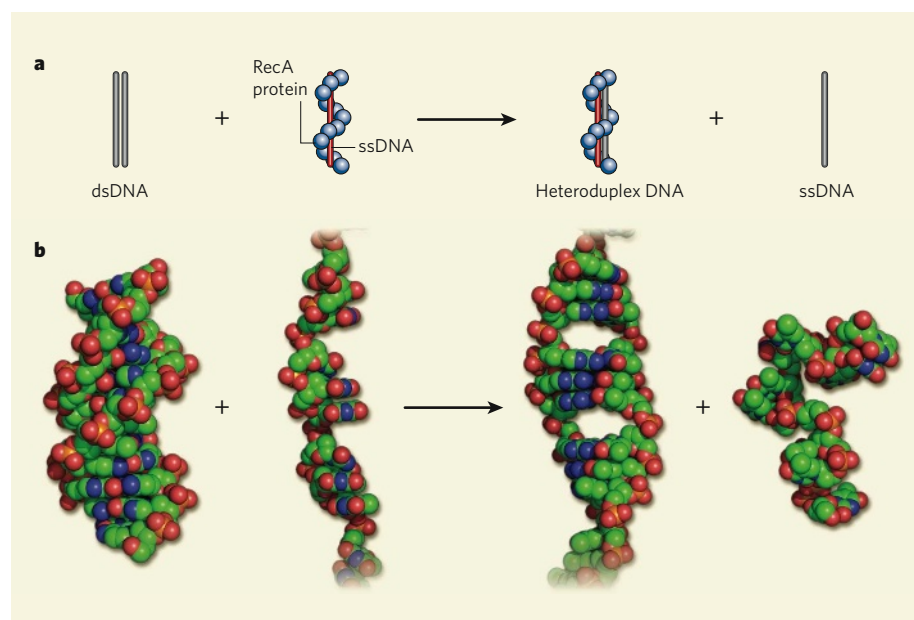
Endocytosis is the process by which cells internalize extracellular material. It is crucial to cell survival and the proper functioning of tissues, being involved in processes as diverse as growth, neural transmission and pathogen clearance. But several opportunistic molecules (bacterial and plant toxins) and pathogens (viruses and bacteria) can exploit the endocytic machinery of a host cell for their own gain<sup>1,2</sup>. Writing in *Science*, Mercer and Helenius<sup>3</sup> report how vaccinia virus (Fig. 1) — a cousin of variola virus, which causes smallpox — deceives host cells into taking it up through endocytosis.

When cells are damaged or dying, for example during programmed cell death (apoptosis), they show several characteristic features. For instance, phosphatidylserine, a lipid that is

abundant in the inner (cytoplasmic) layer of the cell membrane, is redistributed to the outer layer. This phospholipid is thus available to bind to receptors on the surface of phagocytic cells that initiate the apoptotic cell's destruction and engulf it<sup>4</sup>. Another sign of apoptosis is membrane blebbing, or the formation of irregular bulges on the cell membrane. Blebbing also occurs during other processes, such as cell migration and division, but its function is unclear.

When enveloped viruses bud off from their host cell, they inherit a lipid coating (envelope) that has the same composition as the host cell membrane. Mercer and Helenius<sup>3</sup> report that the outer layer of the vaccinia virus envelope contains phosphatidylserine and that this is crucial for infection. They propose that the





**Figure 1 | DNA strand exchange promoted by RecA protein.** **a**, The prototypical DNA-strand-exchange reaction. Double-stranded DNA (dsDNA) pairs with the RecA presynaptic filament, which consists of RecA protein and single-stranded DNA (ssDNA), to produce heteroduplex DNA bound by RecA and the exchanged ssDNA. **b**, Structures of the participating DNA molecules: B-form dsDNA; ssDNA within the presynaptic filament (as determined by Chen and colleagues<sup>1</sup>, protein not shown); dsDNA within the filament<sup>1</sup> (protein not shown); and randomly coiled ssDNA.

species in DNA strand exchange that searches for a homologous sequence within the dsDNA. Once found, DNA strand exchange occurs as a concerted swap of DNA strands. The hydrolysis of ATP inactivates the filament, and permits disassembly of the complexes<sup>2</sup>.

Herein lie the mysteries of DNA strand exchange. How does the RecA nucleoprotein filament recognize DNA sequence identity? And, on finding it, how does the exchange occur? How is the stability of dsDNA overcome? Partial answers to these questions emerged from biochemical studies. The homology search is a 'simple' collisional process because ATP hydrolysis is not essential, only ATP binding. In fact, ATP binding by the RecA nucleoprotein filament is sufficient for DNA strand exchange<sup>3</sup>. The free energy of ATP hydrolysis is not directly involved in the exchange of DNA strands; rather, the free energy of presynaptic-filament binding to the dsDNA 'activates' it by extending and untwisting it, making the duplex DNA a willing participant in the exchange process<sup>4</sup>. ATP hydrolysis then allows dissociation of all participants: a classic case of 'credit card' energetics (expend now, pay later ...).

Structural information derived from electron microscopy was particularly revealing with respect to these questions. The ATP-bound form of the RecA nucleoprotein filament is extended by about 50% relative to standard B-form DNA, with around 6.2 RecA monomers and 18 DNA base pairs per turn<sup>5</sup>; this extended filament is also seen for all RecA homologues<sup>6</sup>. In contrast, the DNA in the inactive ADP-bound nucleoprotein filament is less extended. Thus, the RecA nucleoprotein

filament undergoes ligand-induced structural transitions between an active, extended filament and an inactive, compact filament<sup>5</sup>. The electron microscopy studies also revealed that the RecA nucleoprotein filaments are structurally polymorphic, varying in pitch, width and extension<sup>5</sup>, highlighting the challenge facing higher-resolution structural analysis.

Nonetheless, in 1992 the crystal structure of a RecA filament was solved<sup>7</sup>, but the structure lacked DNA and was of the inactive ADP-bound compact form<sup>8</sup>. For many years, the absence of a structure of the active form confounded molecular and mechanistic interpretation<sup>9</sup>.

However, Chen and colleagues<sup>1</sup> now elucidate structures of RecA assembled on single-stranded DNA (ssDNA) and on dsDNA, defining both the substrate and product forms of the reaction, respectively. How did they succeed? The authors recognized that the intrinsic conformational flexibility of the RecA nucleoprotein filaments and their capacity to self-assemble indefinitely might hinder crystallization in the active state. Their solution to these problems was ingenious, and is applicable to other self-assembling systems.

First, they created 'pre-polymerized' assemblies of RecA protein by fusing four, five or six monomers of RecA into a single polypeptide chain. To prevent indefinite polymerization of the resulting 'mini-filaments', the sites for monomer-monomer interactions were deleted from the first and last monomers in the chain. Despite the many potential pitfalls, this approach worked splendidly, producing functional proteins. When assembled in the presence of an ATP analogue on DNA that exactly accommodated these fusion proteins (15 and



## 50 YEARS AGO

The launching of *Sputnik 3* (Satellite 1958  $\delta$ ) was announced from Moscow on May 15. The satellite was stated to be conical in shape, with a length of 12.3 ft. excluding aerials, a base diameter of 5.7 ft. and a weight of 2,926 lb., including 2,134 lb. of apparatus. The experiments for which the satellite is designed include studies of cosmic rays, geomagnetism, solar radiation and micrometeorites, and the results are to be telemetered back to the Earth. The satellite is equipped with solar batteries and carries a radio transmitter with a frequency of 20.005 Mc./sec. There are two other objects in orbit with the satellite, namely, the rocket which performed the last stage of propulsion and a nose cone which protected the instruments during the climb through the atmosphere.

From *Nature* 24 May 1958.

## 100 YEARS AGO

The last half-yearly number of the *Journal of the Royal Anthropological Institute* contains an important memoir, prepared by two enthusiastic Scotch anthropologists, Messrs. Gray and Tocher, on the pigmentation of hair and eyes among the school children of Scotland ... The highest density of fair hair is to be found in the great river valleys opening on the German Ocean and in the Western Isles. In the former case, this probably points to invasions of a blonde race into those regions. Similarly, the higher percentage of fair hair in the Spey valley and in the Western Isles implies inroads of the Vikings or Norsemen. It is perhaps pushing the evidence too far when the writers suggest that the high percentage of fair-haired girls in the neighbourhood of Dunfermline is due to the train of blonde damsels who are supposed to have accompanied the Saxon princess Margaret, who about the time of the Norman Conquest became Queen of Malcolm Canmore.

From *Nature* 21 May 1908.

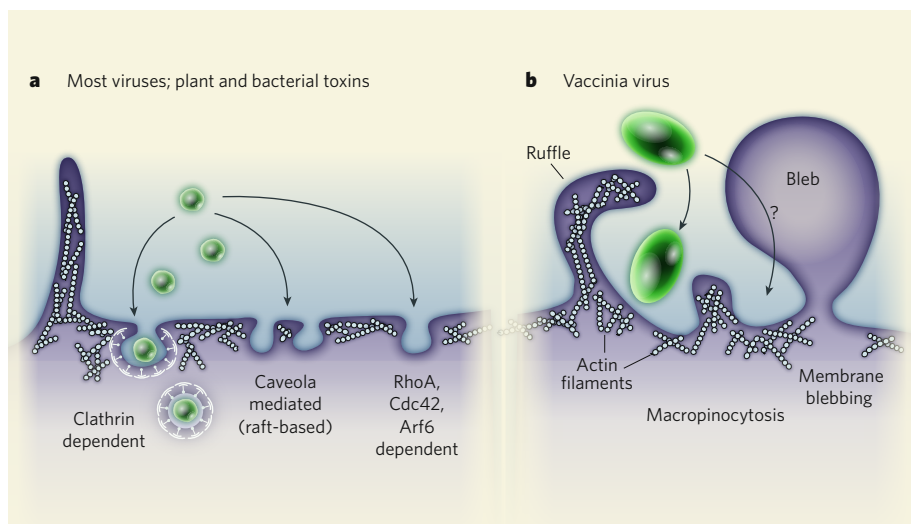
50 & 100 YEARS AGO

underlying mechanism is as follows. When a cell becomes infected with the virus, it displays apoptotic features, including the presence of phosphatidylserine in its outer membrane layer. Thus, when the virus buds off from the cell, it inherits this as part of its envelope. Consequently, cells probably 'mistake' the unusually large vaccinia virus for an apoptotic body (the debris of dying cells) and engulf it.

Mercer and Helenius<sup>3</sup> find that vaccinia virus seems to enter its host cell through an endocytic process called macropinocytosis, which normally mediates fluid uptake. Like virus budding, virus uptake also exploits apoptotic mechanisms. The authors show that vaccinia virus initially binds to cytoplasmic protrusions called filopodia that extend from the surface of the target cell. It moves along them towards the cell body, and then somehow sends signals into the cell, stimulating extensive membrane blebbing. During blebbing, the actin network that forms the scaffolding of the cell beneath the cell membrane becomes detached (Fig. 2). Blebbing is transient, and soon after, the network reassembles in the same cellular location, and the blebs retract<sup>5</sup>.

The authors find that virus internalization by this macropinocytosis-like process is mediated by the blebs, as bleb retraction and re-formation of the actin network coincide with virus entry. Moreover, the drug blebbistatin, which inhibits blebbing<sup>6</sup>, blocks virus entry. In addition, virus internalization requires several proteins (including actin, PAK1, Rac1 and various lipid- and protein-kinase enzymes) that are involved in membrane blebbing<sup>3</sup>. Thus, blebbing might participate in endocytosis, probably when the bleb is retracting and the actin system is re-forming, as the bleb could fold over, or invaginate — a process that would resemble macropinocytosis.

The possibility that blebbing and macropinocytosis are two entirely independent



**Figure 2 | Vaccinia virus chooses to be different.** **a**, Most viruses, as well as plant and bacterial toxins, enter host cells through the classic method of endocytosis, which involves membrane invagination and pinching off of the membrane to form an intracellular transport vesicle. Three different forms of this type of endocytosis are shown. **b**, Mercer and Helenius<sup>3</sup> find that vaccinia virus enters by an endocytic mechanism resembling macropinocytosis. On the cell surface, the virus triggers membrane blebbing, which might also lead to the formation of membrane invaginations that will evolve into transport vesicles. The actin network, which is normally present beneath the cell membrane and is involved in various endocytic processes, is absent in the bleb, but re-forms during bleb retraction.

processes, both of which are stimulated by vaccinia virus, is equally valid. Specifically, blebbistatin inhibits the myosin II protein, which is required for blebbing. But it can also inhibit myosin-II-independent processes such as macropinocytosis<sup>7</sup>. So Mercer and Helenius's observations raise the question of whether macropinocytosis should be subdivided into at least two types: the traditional type in which membrane ruffling (small, dynamic folds of the membrane; Fig. 2b) precedes vesicle formation; and the type that involves blebbing. Multi-modal macropinocytosis would fit well with the increasing number of other types of endocytosis that are being identified<sup>8</sup>.

That opportunistic pathogens exploit various mechanisms for entry and replication within host cells is also documented in a study<sup>9</sup> of the bacterium *Pseudomonas aeruginosa*. This pathogen induces the formation of very large membrane blebs in epithelial cells, entering the blebs and replicating there. The blebs are quite translucent, and do not seem to contain cytoskeletal elements such as actin. Moreover, the bacteria are highly motile within the blebs. But the exact mechanism of bacterial entry into them remains elusive.

Discoveries often raise new questions, and Mercer and Helenius's work<sup>3</sup> is no exception. First, what is the exact relationship between blebbing and macropinocytosis? Cholesterol, for example, is required for both virus infection and macropinocytosis. Is it also required for blebbing? Is virus-induced blebbing cell-type-specific? What happens in polarized cells, in which membrane components and structural elements vary in different parts of the cell, as opposed to the non-polarized cell lines that were studied here? As the active form of

the Arf6 protein inhibits virus infection, one might also wonder how Arf6 is involved in this process. Much is to be learnt about the mechanisms and pathways underlying the internalization of opportunistic pathogens such as vaccinia virus. Ironically, further knowledge about endocytosis itself is likely to come from studies of pathogens.

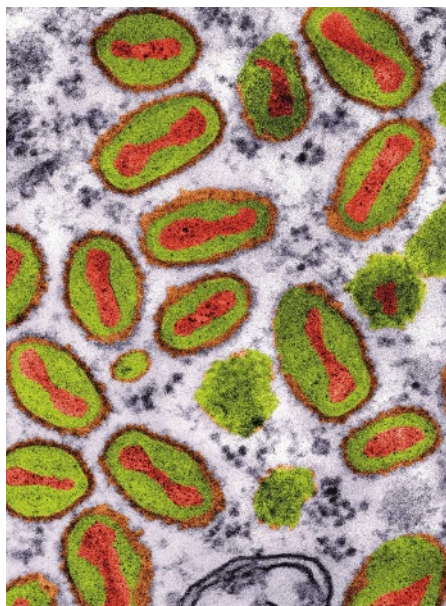
Kirsten Sandvig is in the Centre for Cancer Biomedicine, University of Oslo, and the Department of Biochemistry, Institute for Cancer Research, Norwegian Radium Hospital, Rikshospitalet University Hospital, Montebello, 0310 Oslo, Norway. Bo van Deurs is at the Institute of Cellular and Molecular Medicine, The Panum Building, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark.  
e-mails: ksandvig@radium.uio.no;  
bvd@sund.ku.dk

1. Marsh, M. & Helenius, A. *Cell* **124**, 729–740 (2006).
2. Sandvig, K. & van Deurs, B. *Gene Ther.* **12**, 865–872 (2005).
3. Mercer, J. & Helenius, A. *Science* **320**, 531–535 (2008).
4. Henson, P. M., Bratton, D. L. & Fadok, V. A. *Curr. Biol.* **11**, R795–R805 (2001).
5. Charras, G. T., Hu, C.-K., Coughlin, M. & Mitchison, T. J. *J. Cell Biol.* **175**, 477–490 (2006).
6. Straight, A. F. et al. *Science* **299**, 1743–1747 (2003).
7. Shu, S., Liu, X. & Korn, E. D. *Proc. Natl Acad. Sci. USA* **102**, 1472–1477 (2005).
8. Sandvig, K., Torgersen, M. L., Raa, H. A. & van Deurs, B. *Histochem. Cell Biol.* **129**, 267–276 (2008).
9. Angus, A. A. et al. *Infect. Immun.* **76**, 1992–2001 (2008).

#### Correction

In the News & Views article "Palaeoclimate: Windows on the greenhouse" by Ed Brook (*Nature* **453**, 291–292; 2008), the wrong credit was used for the picture on page 291. The figure in fact came from the 1978 doctoral thesis of W. Berner (University of Bern).

K. BOLLER/SPL



**Figure 1 | Source of infection.** False-colour electron microscopy image of vaccinia virus.



# An extremely luminous X-ray outburst at the birth of a supernova

A. M. Soderberg<sup>1,2</sup>, E. Berger<sup>1,2</sup>, K. L. Page<sup>3</sup>, P. Schady<sup>4</sup>, J. Parrent<sup>5</sup>, D. Pooley<sup>6</sup>, X.-Y. Wang<sup>7</sup>, E. O. Ofek<sup>8</sup>, A. Cucchiara<sup>9</sup>, A. Rau<sup>8</sup>, E. Waxman<sup>10</sup>, J. D. Simon<sup>8</sup>, D. C.-J. Bock<sup>11</sup>, P. A. Milne<sup>12</sup>, M. J. Page<sup>4</sup>, J. C. Barentine<sup>13</sup>, S. D. Barthelmy<sup>14</sup>, A. P. Beardmore<sup>3</sup>, M. F. Bietenholz<sup>15,16</sup>, P. Brown<sup>9</sup>, A. Burrows<sup>1</sup>, D. N. Burrows<sup>9</sup>, G. Byrneson<sup>17</sup>, S. B. Cenko<sup>18</sup>, P. Chandra<sup>19</sup>, J. R. Cummings<sup>20</sup>, D. B. Fox<sup>9</sup>, A. Gal-Yam<sup>10</sup>, N. Gehrels<sup>20</sup>, S. Immler<sup>20</sup>, M. Kasliwal<sup>8</sup>, A. K. H. Kong<sup>21</sup>, H. A. Krimm<sup>20,22</sup>, S. R. Kulkarni<sup>8</sup>, T. J. Maccarone<sup>23</sup>, P. Mészáros<sup>9</sup>, E. Nakar<sup>24</sup>, P. T. O'Brien<sup>3</sup>, R. A. Overzier<sup>25</sup>, M. de Pasquale<sup>4</sup>, J. Racusin<sup>9</sup>, N. Rea<sup>26</sup> & D. G. York<sup>27</sup>

**Massive stars end their short lives in spectacular explosions—supernovae—that synthesize new elements and drive galaxy evolution. Historically, supernovae were discovered mainly through their ‘delayed’ optical light (some days after the burst of neutrinos that marks the actual event), preventing observations in the first moments following the explosion. As a result, the progenitors of some supernovae and the events leading up to their violent demise remain intensely debated. Here we report the serendipitous discovery of a supernova at the time of the explosion, marked by an extremely luminous X-ray outburst. We attribute the outburst to the ‘break-out’ of the supernova shock wave from the progenitor star, and show that the inferred rate of such events agrees with that of all core-collapse supernovae. We predict that future wide-field X-ray surveys will catch each year hundreds of supernovae in the act of exploding.**

Stars more massive than about eight times the mass of the Sun meet their death in cataclysmic explosions termed supernovae. These explosions give birth to the most extreme compact objects—neutron stars and black holes—and enrich their environments with heavy elements. It is generally accepted that supernovae are triggered when the stellar core runs out of fuel for nuclear burning and thus collapses under its own gravity (see ref. 1 and references therein). As the collapsing core rebounds, it generates a shock wave that propagates through, and explodes, the star.

The resulting explosion ejects several solar masses of stellar material with a mean velocity<sup>2</sup> of about  $10^4 \text{ km s}^{-1}$ , or a kinetic energy of about  $10^{51} \text{ erg}$ . Less than a solar mass of  $^{56}\text{Ni}$  is synthesized in the explosion, but its subsequent radioactive decay powers<sup>1</sup> the luminous optical light observed to peak 1–3 weeks after the explosion. It is through this delayed signature that supernovae have been discovered both historically and in modern searches.

Although the general picture of core collapse has been recognized for many years, the details of the explosion remain unclear and most supernova simulations fail to produce an explosion. The gaps in our understanding are due to the absence of detailed observations in the first days after the explosion, and the related difficulty in detecting the weak neutrino<sup>3</sup> and gravitational wave signatures of the explosion.

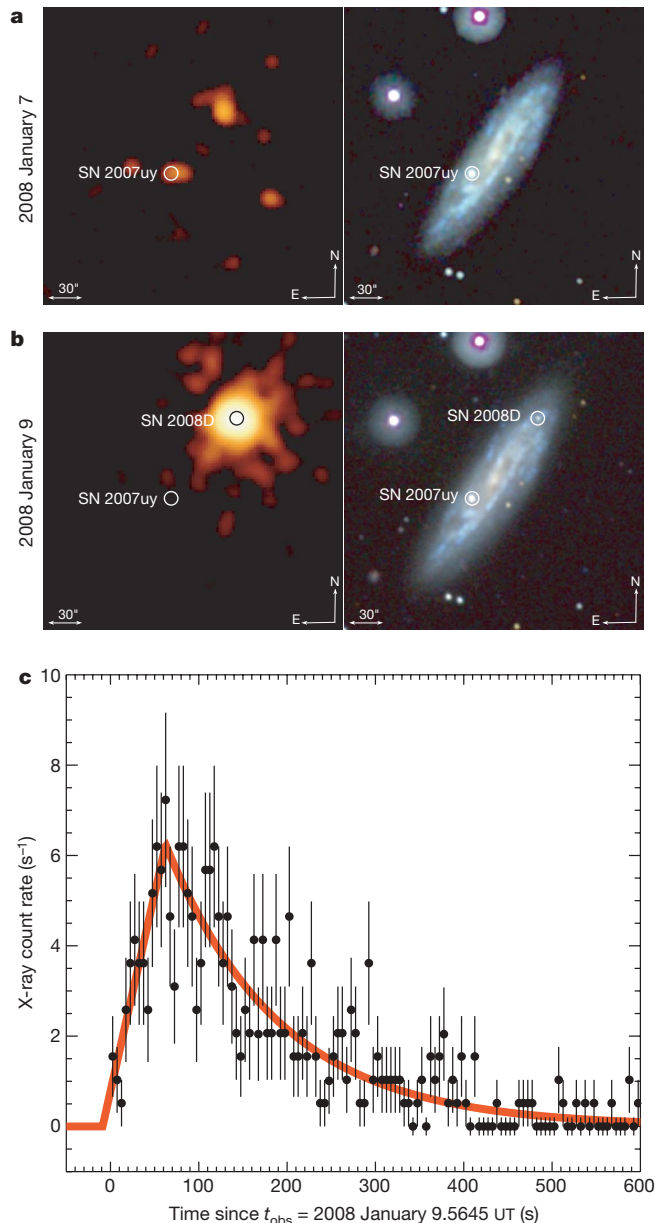
These signals offer a direct view of the explosion mechanism but require the discovery of supernovae at the time of explosion.

In this Article we describe our serendipitous discovery of an extremely luminous X-ray outburst that marks the birth of a supernova of type Ibc. Prompt bursts of X-ray and/or ultraviolet emission have been theorized<sup>4,5</sup> to accompany the break-out of the supernova shock wave through the stellar surface, but their short durations (just seconds to hours) and the lack of sensitive wide-field X-ray and ultraviolet searches have prevented their discovery until now.

Our detection enables an unprecedented early and detailed view of the supernova, allowing us to infer<sup>6</sup> the radius of the progenitor star, its mass loss in the final hours before the explosion, and the speed of the shock as it explodes the star. Drawing on optical, ultraviolet, radio and X-ray observations, we show that the progenitor was compact (radius  $R_* \approx 10^{11} \text{ cm}$ ) and stripped of its outer hydrogen envelope by a strong and steady stellar wind. These properties are consistent<sup>7</sup> with those of Wolf-Rayet stars, the favoured<sup>8</sup> progenitors of type Ibc supernovae.

Wolf-Rayet stars are also argued<sup>9</sup> to give rise to  $\gamma$ -ray bursts (GRBs), a related but rare class of explosions characterized by highly collimated relativistic jets. Our observations, however, indicate an ordinary spherical and non-relativistic explosion and we firmly rule out a GRB connection.

<sup>1</sup>Department of Astrophysical Sciences, Princeton University, Ivy Lane, Princeton, New Jersey 08544, USA. <sup>2</sup>Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA. <sup>3</sup>Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. <sup>4</sup>Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK. <sup>5</sup>Physics and Astronomy Department, Dartmouth College, Hanover, New Hampshire 03755, USA. <sup>6</sup>Astronomy Department, University of Wisconsin, 475 North Charter Street, Madison, Wisconsin 53706, USA. <sup>7</sup>Department of Astronomy, Nanjing University, Nanjing 210093, China. <sup>8</sup>Department of Astronomy, 105-24, California Institute of Technology, Pasadena, California 91125, USA. <sup>9</sup>Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>10</sup>Faculty of Physics, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>11</sup>Radio Astronomy Laboratory, University of California, Berkeley, California 94720, USA. <sup>12</sup>Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. <sup>13</sup>Department of Astronomy, University of Texas at Austin, Austin, Texas 78712, USA. <sup>14</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. <sup>15</sup>Department of Physics and Astronomy, York University, Toronto, Ontario M3J 1P3, Canada. <sup>16</sup>Hartebeestehoeck Radio Observatory, PO Box 443, Krugersdorp, 1740, South Africa. <sup>17</sup>Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634, USA. <sup>18</sup>Space Radiation Laboratory, 220-47, California Institute of Technology, Pasadena, California 91125, USA. <sup>19</sup>Department of Astronomy, University of Virginia, PO Box 400325, Charlottesville, Virginia 22904, USA. <sup>20</sup>CRESST and NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. <sup>21</sup>Institute of Astronomy and Department of Physics, National Tsing Hua University, Hsinchu, Taiwan. <sup>22</sup>Universities Space Research Association, 10211 Wincopin Circle, #500, Columbia, Maryland 21044, USA. <sup>23</sup>School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK. <sup>24</sup>Theoretical Astrophysics, 130-33, California Institute of Technology, Pasadena, California 91125, USA. <sup>25</sup>Max-Planck-Institut für Astrophysik, D-85748 Garching, Germany. <sup>26</sup>University of Amsterdam, Astronomical Institute ‘Anton Pannekoek’, Kruislaan 403, 1098SJ, Amsterdam, The Netherlands. <sup>27</sup>Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637, USA.



**Figure 1 | Discovery image and X-ray light curve of XRO 080109/SN 2008D.** **a**, X-ray (left) and ultraviolet (right) images of the field obtained on 2008 January 7 UT during Swift observations of the type Ibc supernova 2007uy. No source is detected at the position of SN 2008D to a limit of  $\lesssim 10^{-3}$  counts s<sup>-1</sup> in the X-ray band and  $U \gtrsim 20.3$  mag. **b**, Repeated ultraviolet and X-ray observations of the field from January 9 UT during which we serendipitously discovered XRO 080109 and its ultraviolet counterpart. The position of XRO 080109 is right ascension  $\alpha = 09^{\text{h}} 09^{\text{m}} 30.70^{\text{s}}$ , declination  $\delta = 33^{\circ} 08' 19.1''$  (J2000) ( $\pm 3.5''$ ), about 9 kpc from the centre of NGC 2770. **c**, X-ray light curve of XRO 080109 in the 0.3–10 keV band. The data were accumulated in the photon counting mode and were processed using version 2.8 of the Swift software package, including the most recent calibration and exposure maps. The high count rate resulted in photon pile-up, which we correct for by fitting a King function profile to the point spread function (PSF) to determine the radial point at which the measured PSF deviates from the model. The counts were extracted using an annular aperture that excluded the affected 4 pixel core of the PSF, and the count rate was corrected according to the model. Error bars,  $\pm 1\sigma$ . Using a fast rise, exponential decay model (red curve), we determine the properties of the outburst, in particular its onset time,  $t_0$ , which corresponds to the explosion time of SN 2008D. The best-fit parameters are a peak time of  $63 \pm 7$  s after the beginning of the observation, an e-folding time of  $129 \pm 6$  s, and peak count rate of  $6.2 \pm 0.4$  counts s<sup>-1</sup> (90% confidence level using Cash statistics). The best-fit value of  $t_0$  is January 9 13:32:40 UT (that is, 9 s before the start of the observation) with a 90% uncertainty range of 13:32:20 to 13:32:48 UT.

Most importantly, the inferred rate of X-ray outbursts indicates that all core-collapse supernovae produce detectable shock break-out emission. Thus, we predict that future wide-field X-ray surveys will uncover hundreds of supernovae each year at the time of explosion, providing the long-awaited temporal and positional triggers for neutrino and gravitational wave searches.

### Discovery of the X-ray outburst

On 2008 January 9 at 13:32:49 UT, we serendipitously discovered an extremely bright X-ray transient during a scheduled Swift X-ray Telescope (XRT) observation of the galaxy NGC 2770 (distance  $d = 27$  Mpc). Previous XRT observations of the field just two days earlier revealed no pre-existing source at this location. The transient, hereafter designated as X-ray outburst (XRO) 080109, lasted about 400 s, and was coincident with one of the galaxy's spiral arms (Fig. 1). From observations described below, we determine that XRO 080109 is indeed located in NGC 2770, and we thus adopt this association from here on.

The temporal evolution is characterized by a fast rise and exponential decay, often observed for a variety of X-ray flare phenomena (Fig. 1). We determine the onset of the X-ray emission to be  $9^{+20}_{-8}$  s before the beginning of the observation, implying an outburst start time ( $t_0$ ) of January 9.5644 UT. The X-ray spectrum is best fitted by a power law ( $N(E) \propto E^{-\Gamma}$ , where  $N$  and  $E$  are the photon number and energy, respectively) with a photon index of  $\Gamma = 2.3 \pm 0.3$ , and a hydrogen column density of  $N_{\text{H}} = 6.9^{+1.8}_{-1.5} \times 10^{21}$  cm<sup>-2</sup>, in excess of the absorption within the Milky Way (see Supplementary Information). The inferred unabsorbed peak flux is  $F_{\text{X,p}} \approx 6.9 \times 10^{-10}$  erg cm<sup>-2</sup> s<sup>-1</sup> (0.3–10 keV). We also measure significant spectral softening during the outburst.

The XRO was in the field of view of the Swift Burst Alert Telescope (BAT; 15–150 keV) beginning 30 min before and continuing throughout the outburst, but no  $\gamma$ -ray counterpart was detected. Thus, the outburst was not a GRB (see also Supplementary Information). Integrating over the duration of the outburst, we place a limit on the  $\gamma$ -ray fluence of  $f_{\gamma} \lesssim 8 \times 10^{-8}$  erg cm<sup>-2</sup> ( $3\sigma$ ), a factor of three times higher than an extrapolation of the X-ray spectrum to the BAT energy band.

The total energy of the outburst is thus  $E_{\text{X}} \approx 2 \times 10^{46}$  erg, at least three orders of magnitude lower<sup>10</sup> than GRBs. The peak luminosity is  $L_{\text{X,p}} \approx 6.1 \times 10^{43}$  erg s<sup>-1</sup>, several orders of magnitude larger than the Eddington luminosity (the maximum luminosity for a spherically accreting source) of a solar mass object, outbursts from ultra-luminous X-ray sources and type I X-ray bursts. In summary, the properties of XRO 080109 are distinct from those of all known X-ray transients.

### The birth of a supernova

Simultaneous observations of the field with the co-aligned Ultraviolet/Optical Telescope (UVOT) on board Swift showed no evidence for a contemporaneous counterpart. However, UVOT observations just 1.4 h after the outburst revealed<sup>11</sup> a brightening ultraviolet/optical counterpart. Subsequent ground-based optical observations also uncovered<sup>11–13</sup> a coincident source.

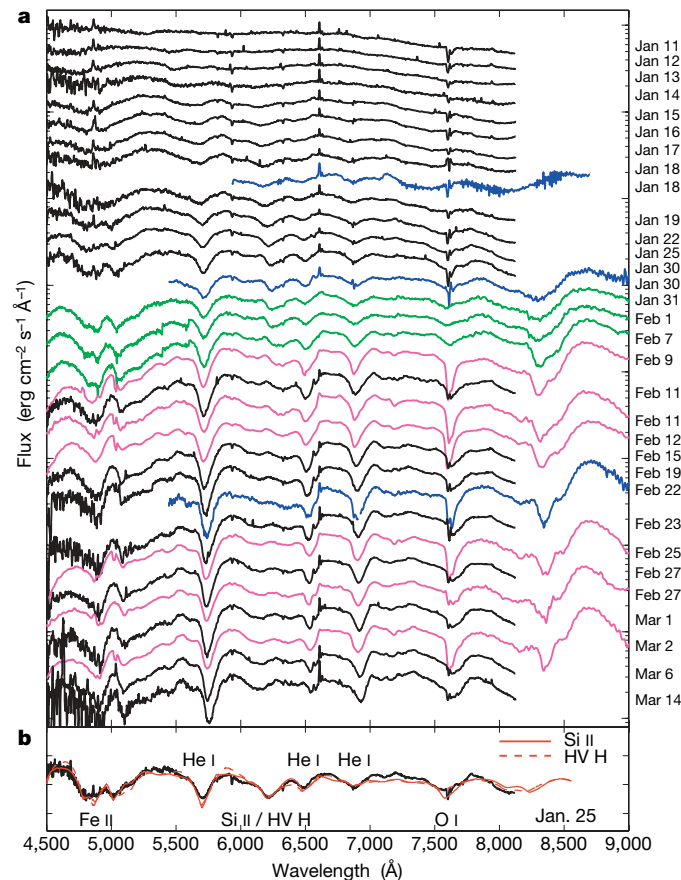
We promptly obtained optical spectroscopy of the counterpart with the Gemini North 8-m telescope beginning 1.74 d after the outburst (Fig. 2). The spectrum is characterized by a smooth continuum with narrow absorption lines of Na I (wavelengths 5,890 and 5,896 Å) at the redshift of NGC 2770. More importantly, we note broad absorption features near 5,200 and 5,700 Å and a drop-off beyond 7,000 Å, strongly suggestive of a young supernova. Subsequent observations confirmed these spectral characteristics<sup>11,14</sup>, and the transient was classified<sup>11,15</sup> as type Ibc SN 2008D based on the lack of hydrogen and weak silicon features.

Thanks to the prompt X-ray discovery, the temporal coverage of our optical spectra exceeds those of most supernovae, rivalling even the best-studied GRB-associated supernovae, and SN 1987A (Fig. 2). We see a clear evolution from a mostly featureless continuum to broad absorption lines, and finally to strong absorption features with moderate widths. Moreover, our spectra reveal the emergence of



strong He I features within a few days of the outburst (see also ref. 16). Thus, SN 2008D is a He-rich type Ibc supernova, unlike<sup>17</sup> GRB-associated supernovae. Observations at high spectral resolution further reveal significant host galaxy extinction, with  $A_V \approx 1.2$ – $2.5$  mag (see Supplementary Information).

The well-sampled ultraviolet/optical light curves in ten broadband filters (2,000–10,000 Å) exhibit two distinct components (Fig. 3). First, an ultraviolet-dominated component that peaks about a day after the X-ray outburst, and which is similar to very early observations<sup>18</sup> of the GRB-associated SN 2006aj. The second component is significantly redder and peaks on a timescale of about 20 d, consistent with observations of all type Ibc supernovae. Accounting for an extinction of  $A_V = 1.9$  mag (Fig. 3), the absolute peak brightness of the second component is  $M_V \approx -16.7$  mag, at the low end of the distribution<sup>19</sup> for type Ibc supernovae and GRB-associated supernovae.

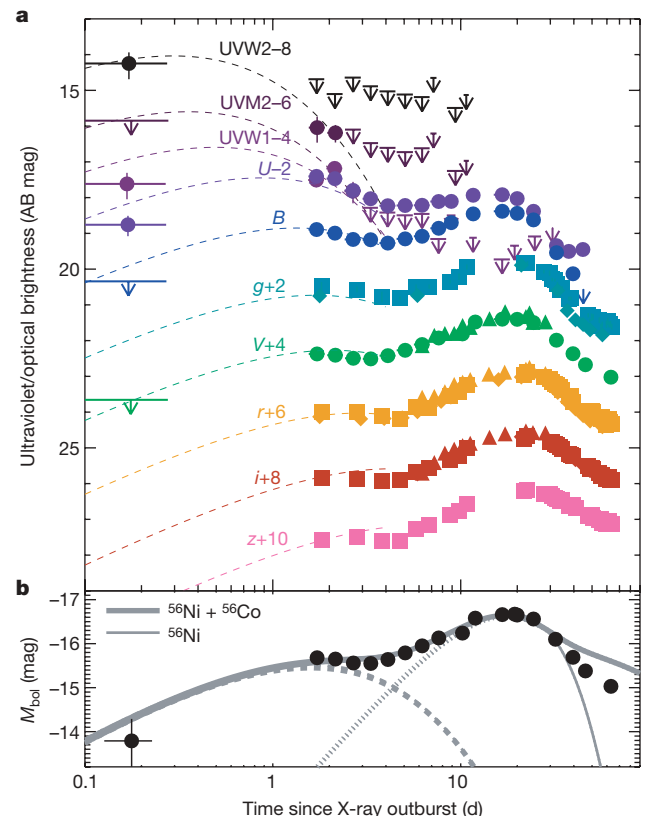


**Figure 2 | Optical spectra of XRO 080109/SN 2008D, and model fit. a,** The spectra are plotted logarithmically in flux units and shifted for clarity. **b,** A model fit to the January 25 UT spectrum using the spectral fitting code SYNOW. We identify several strong features attributed to He I, O I and Fe II, indicating a type Ibc classification. In addition, we find an absorption feature at 6,200 Å that can be identified as Si II or high velocity H I (HV H; see Supplementary Information for details). The observations were performed using the following facilities: The Gemini Multi-Object Spectrograph (GMOS) on the Gemini North 8-m telescope (black); the Dual Imaging Spectrograph (DIS) on the Apache Point 3.5-m telescope (blue); the Double Spectrograph (DBSP) on the Palomar Hale 200-inch telescope (green); and the Low Resolution Spectrograph (LRS) on the Hobby-Eberly 9.2-m telescope (magenta). The details of the observational set-up and the exposure times are provided in Supplementary Information. The data were reduced using the gemini package within the Image Reduction and Analysis Facility (IRAF) software for the GMOS data. All other observations were reduced using standard packages in IRAF. The supernova spectra were extracted from the two-dimensional data using a nearby background region to reduce the contamination from host galaxy emission. Absolute flux calibration was achieved using observations of the standard stars Feige 34 and G191B2B.

### A shock break-out origin

As some type Ibc supernovae harbour GRBs, we investigate the possibility that the XRO is produced by a relativistic outflow. In this scenario, the X-ray flux and simultaneous upper limits in the ultraviolet/optical require the outflow to be ultra-relativistic with a bulk Lorentz factor  $\gamma \approx 90$ , but its radius to be only  $R \approx 10^{10}$  cm; here  $\gamma \equiv (1 - \beta^2)^{-1/2}$  and  $\beta \equiv v/c$ , where  $v$  is the outflow velocity and  $c$  is the speed of light. However, given the observed duration of the outburst, we expect<sup>20</sup>  $R \approx 4\gamma^2 ct \approx 10^{17}$  cm, indicating that the relativistic outflow scenario is not self-consistent (see Supplementary Information for details).

We are left with a trans- or non-relativistic origin for the outburst, and we consider supernova shock break-out as a natural scenario. The break-out is defined by the transition from a radiation-mediated to a collisional (or collisionless<sup>21</sup>) shock as the optical depth of the outflow decreases to unity. Such a transition has long been predicted<sup>4,5</sup> to produce strong, thermal ultraviolet/X-ray emission at



**Figure 3 | Optical and ultraviolet light curves of XRO 080109/SN 2008D, and model fit. a,** Optical and ultraviolet light curves. Data are from Swift UVOT (circles), Palomar 60-inch telescope (squares), Gemini/GMOS (diamonds), and the SLOIS telescope (triangles). Tables summarizing the observations and data analysis are available in Supplementary Information. The data have not been corrected for host galaxy extinction and have been offset (as labelled) for clarity. We fit the data before 3 d with a cooling envelope blackbody emission model<sup>6</sup> (dashed lines) that accounts for host extinction ( $A_V$ ). We find a reasonable fit to the data with  $R_* \approx 10^{11}$  cm,  $E_K \approx 2 \times 10^{51}$  erg,  $M_{ej} \approx 5 M_\odot$  and  $A_V \approx 1.9$  mag, consistent with the constraints from the high-resolution optical spectrum. The radius and temperature of the photosphere at 1 d are  $R_{ph} \approx 3 \times 10^{14}$  cm and  $T_{ph} \approx 10^4$  K, respectively. Error bars are  $1\sigma$ ; down-pointing arrows are upper limits ( $3\sigma$ ). **b,** The absolute bolometric magnitude light curve (corrected for host extinction). The dashed lines are the same cooling envelope model described above, while the short-dashed lines are models of supernova emission powered by radioactive decay. The solid lines are combined models taking into account the decay of  $^{56}\text{Ni}$  (thin line) and  $^{56}\text{Ni} + ^{56}\text{Co}$  (thick line). The supernova models provide an independent measure of  $E_K$  and  $M_{ej}$ , as well as  $M_{Ni}$  (see Supplementary Information for a detailed discussion of the models). We find values that are consistent to within 30% with those inferred from the cooling envelope model.

the time of explosion. A non-thermal component at higher energies may be produced<sup>22</sup> by multiple scatterings of the photons between the ejecta and a dense circumstellar medium (bulk comptonization).

We attribute the observed non-thermal outburst to comptonized emission from shock break-out, indicating that the associated thermal component must lie below the XRT low energy cut-off,  $\sim 0.1$  keV. With the reasonable assumption that the energy in the thermal ( $E_{\text{th}}$ ) and comptonized components is comparable, we constrain<sup>6</sup> the radius at which shock break-out occurs to  $R_{\text{sbo}} \gtrsim 7 \times 10^{11} (T)^{-4/7} (E_X)^{3/7}$  cm (here  $T$  is in units of 0.1 keV, and  $E_X$  in units of  $2 \times 10^{46}$  erg). This is consistent with a simple estimate derived from the rise time of the outburst,  $R_{\text{sbo}} = c\delta t \approx 10^{12}$  cm, and larger than the typical radii of Wolf-Rayet stars<sup>23</sup>,  $\sim 10^{11}$  cm. We therefore attribute the delayed shock break-out to the presence of a dense stellar wind, similar<sup>6,18</sup> to the case of the GRB-associated supernova SN 2006aj.

The shock velocity at break-out is<sup>6</sup>  $(\gamma\beta) \lesssim 1.1$  and the outflow is thus trans-relativistic, as expected<sup>24</sup> for a compact progenitor. Using these constraints, the inferred optical depth of the ejecta to thermal X-rays is  $\tau_{\text{ej}} \approx 1.5 (E_X) (R_{\text{sbo}})^{-2} (\gamma - 1)^{-1} \approx 3$  (here  $E_X$  is normalized as above, and  $R_{\text{sbo}}$  is in units of  $7 \times 10^{11}$  cm), and comptonization is thus efficient, confirming our model. Equally important, as the ejecta expand outward the optical depth of the stellar wind decreases and the spectrum of the comptonized emission is expected<sup>22</sup> to soften, in agreement with the observed trend.

The shock break-out emission traces the wind mass-loss rate of the progenitor,  $\dot{M}$ , in the final hours leading up to the explosion. The inferred density indicates  $\dot{M} \approx 4\pi v_w R_{\text{sbo}} / \kappa \approx 10^{-5} M_{\odot} \text{ yr}^{-1}$ ; here  $\kappa \approx 0.4 \text{ cm}^2 \text{ g}^{-1}$  is the Thomson opacity for an ionized hydrogen wind and  $v_w \approx 10^3 \text{ cm s}^{-1}$  is the typical<sup>7</sup> wind velocity for Wolf-Rayet stars. The mass-loss rate is consistent<sup>7</sup> with the average values inferred for Galactic Wolf-Rayet stars, and, along with the inferred compact stellar radius and the lack of hydrogen features, leads us to conclude that the progenitor was a Wolf-Rayet star.

## Two ultraviolet/optical emission components

The early ultraviolet/optical emission ( $t \lesssim 3$  d, where  $t$  is time since  $t_0$ ) appears to be a distinct component, based on its different temporal behaviour and bluer colours (Fig. 3). We attribute this early emission to cooling of the outer stellar envelope following the passage of the shock through the star and its subsequent break-out (marked by the X-ray outburst). The expected blackbody radiation is characterized<sup>6</sup> by the photospheric radius and temperature, which evolve with  $t$  respectively as  $R_{\text{ph}} \propto t^{0.8}$  and  $T_{\text{ph}} \propto t^{-0.5}$ , and depend on the total ejecta kinetic energy ( $E_K$ ) and mass ( $M_{\text{ej}}$ ), and on the stellar radius before the explosion ( $R_*$ ).

The model light curves provide a good fit to the early ultraviolet/optical data (Fig. 3). The implied stellar radius is  $R_* \approx 7 \times 10^{10}$  cm, consistent with that expected<sup>23</sup> for a Wolf-Rayet progenitor. Moreover, this value is smaller than the shock break-out radius, confirming our earlier inference that the break-out occurs in the extended stellar wind.

The ratio of  $E_K$  and  $M_{\text{ej}}$  also determines the shape of the main supernova light curve (see, for example, ref. 25), and the mass of  $^{56}\text{Ni}$  synthesized in the explosion ( $M_{\text{Ni}}$ ) determines<sup>26</sup> its peak optical luminosity. To break the degeneracy between  $E_K$  and  $M_{\text{ej}}$ , we measure the photospheric velocity from the optical spectra at maximum light,  $v_{\text{ph}} = 0.3 (E_K/M_{\text{ej}})^{1/2} \approx 11,500 \text{ km s}^{-1}$ ; this is comparable to that of ordinary type Ibc supernovae, but somewhat slower<sup>17</sup> than GRB-associated supernovae (Fig. 2 and Supplementary Information). We find that both light curve components are self-consistently fitted with  $E_K \approx (2-4) \times 10^{51} \text{ erg}$ ,  $M_{\text{ej}} \approx 3-5 M_{\odot}$ , and  $M_{\text{Ni}} \approx 0.05-0.1 M_{\odot}$  (Fig. 3).

## Long-lived X-ray and radio emission

Whereas ultraviolet/optical observations probe the bulk material, radio and X-ray emission trace fast ejecta. Our Swift follow-up observations of the XRO revealed fainter X-ray emission several hours after

the explosion, with  $L_X \approx 2 \times 10^{40} \text{ erg s}^{-1}$  ( $t \approx 0.2$  d). This emission exceeds the extrapolation of the outburst by many orders of magnitude, indicating that it is powered by a different mechanism. Using a high-angular-resolution observation from the Chandra X-ray Observatory on January 19.86 UT, we detect the supernova with a luminosity  $L_X = (1.0 \pm 0.3) \times 10^{39} \text{ erg s}^{-1}$  (0.3–10 keV), and further resolve three nearby sources contained within the 18-arcsec resolution element of XRT. Correcting all XRT observations for these sources, we find that the long-lived X-ray emission decays steadily as  $F_X \propto t^{-0.7}$  (Supplementary Information).

Using the Very Large Array (VLA) on January 12.54 UT, we further discovered a new radio source at the position of the supernova that was not present on January 7 UT. Follow-up observations were obtained at multiple frequencies between 1.4 and 95 GHz using the VLA, the Combined Array for Research in Millimeter-wave Astronomy (CARMA) and the Very Long Baseline Array (VLBA).

The broadband radio emission on January 14 UT reveals a spectral peak,  $\nu_p \approx 43$  GHz, with a flux density,  $F_{\nu,p} \approx 4 \text{ mJy}$ , and a low frequency spectrum,  $F_{\nu} \propto \nu^{-2.5}$ . Subsequent observations show that  $\nu_p$  cascades to lower frequencies, similar to the evolution observed in other type Ibc supernovae (see, for example, ref. 27). The passage of  $\nu_p$  through each frequency produces a light curve peak. We measure  $F_{\nu} \propto t^{1.4}$  and  $F_{\nu} \propto t^{-1.2}$  for the light curve rise and decline, respectively (Fig. 4).

We note that our X-ray and radio observations of SN 2008D are the earliest ever obtained for a normal type Ibc supernova. At  $t \approx 10$  d, the X-ray and peak radio luminosities are several orders of magnitude lower<sup>28,29</sup> than those of GRB afterglows but comparable<sup>30,31</sup> to those of normal type Ibc supernovae.

## The properties of the fast ejecta

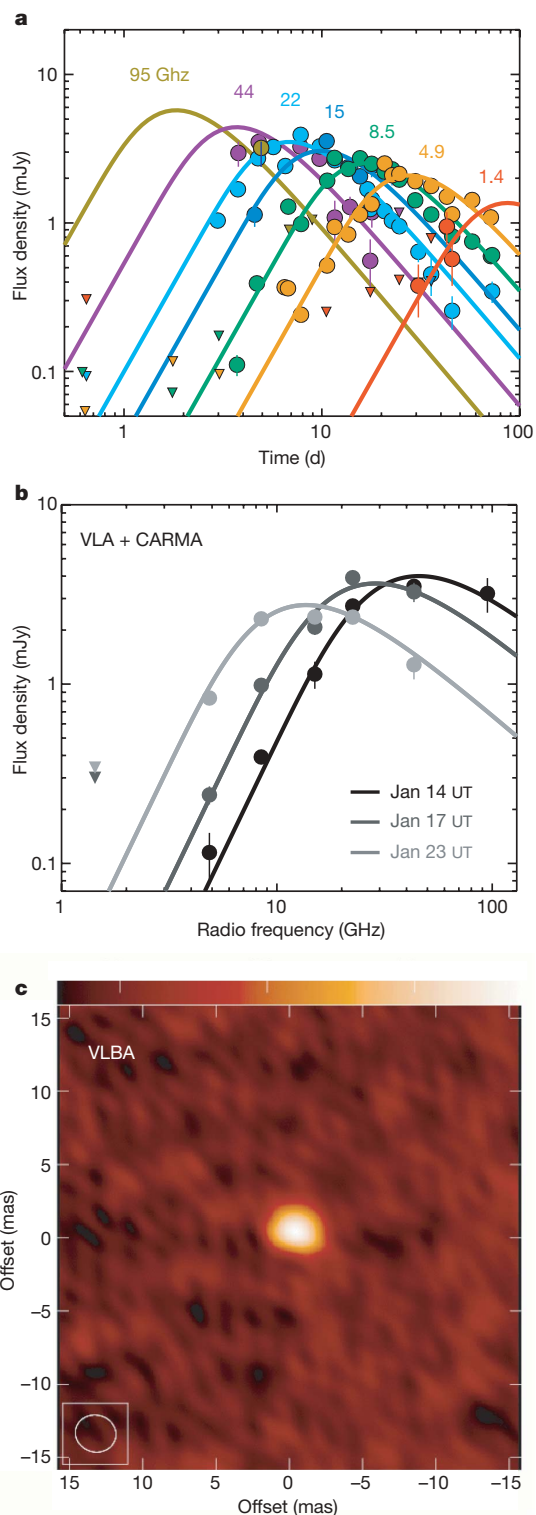
Radio synchrotron emission is produced<sup>32</sup> by relativistic electrons accelerated in the supernova shock as they gyrate in the amplified magnetic field. Self-absorption suppresses the spectrum below the peak to  $F_{\nu} \propto \nu^{2.5}$ , in excellent agreement with our observations. In this context, we infer<sup>33,34</sup> the radius of the fast ejecta, using the measured  $\nu_p$  and  $L_{\nu,p}$ , to be  $R \approx 3 \times 10^{15}$  cm at  $t \approx 5$  d. The implied mean velocity is  $\beta \approx 0.25$ , clearly ruling out relativistic ejecta.

With this conclusion there are two possibilities for the ejecta dynamics. First, the supernova may be in free expansion,  $R \propto t$ , consistent with observations of type Ibc supernovae (see, for example, ref. 27). Alternatively, the ejecta may have been relativistic at early time and then rapidly decelerated, leading to  $R \propto t^{2/3}$ . In the latter scenario, the dynamics are governed<sup>35</sup> by the Sedov-Taylor solution. As discussed in Supplementary Information, the temporal evolution of the radio light curves is clearly inconsistent with the Sedov-Taylor model, ruling out even early relativistic expansion.

Thus, the radio emission is produced by freely expanding ejecta, indicative of the broad velocity structure expected<sup>24</sup> for ordinary core-collapse supernovae. The standard formulation<sup>27</sup> provides an excellent fit to the data (Fig. 4) and indicates that the energy coupled to fast material is  $E_{K,R} \approx 10^{48} \text{ erg}$  (here subscript K,R indicates kinetic energy probed by radio observations), just 0.1% of the total kinetic energy. Moreover, the inferred density profile is  $\rho(r) \propto r^{-2}$  (where  $r$  is the radius from the explosion site), as expected for a steady stellar wind. The inferred mass-loss rate,  $\dot{M} \approx 7 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$ , is in agreement with our shock break-out value, indicating a stable mass loss rate in the final  $\sim 3$  yr to  $\sim 3$  h of the progenitor's life.

The radio-emitting electrons also account for the late X-ray emission through their inverse Compton (IC) upscattering of the supernova optical photons (with a luminosity  $L_{\text{opt}}$ ). The expected<sup>6</sup> X-ray luminosity is  $L_{\text{IC}} \approx 3 \times 10^{39} (E_{K,R}) (L_{\text{opt}}) (t)^{-2/3} \text{ erg s}^{-1}$  (where  $E_{K,R}$  is in units of  $10^{48} \text{ erg}$ ,  $L_{\text{opt}}$  in  $10^{42} \text{ erg s}^{-1}$ , and  $t$  in days), in excellent agreement with the observations by XRT and the Chandra X-ray Observatory. We note that the synchrotron contribution in the X-ray band is lower by at least two orders of magnitude.

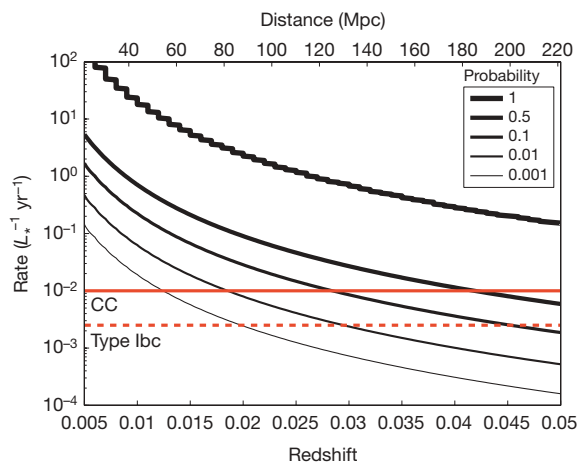




**Figure 4 | Radio light curves, spectra and image of XRO 080109/SN 2008D.** Radio data from 1.4 to 95 GHz were obtained with the VLA, CARMA and the VLBA (circles are detections and inverted triangles represent  $3\sigma$  upper limits). Error bars are  $1\sigma$ . The flux measurements and a description of the data analysis are provided in Supplementary Information. **a**, Radio light curves with a model of synchrotron self-absorbed emission arising<sup>27</sup> from shocked material surrounding the freely expanding supernova. We adopt a shock compression factor of  $\eta = 4$  for the post-shock material and assume that the electrons and magnetic fields each contribute 10% to the total post-shock energy density. The best-fit model (solid lines) implies the following physical parameters and temporal evolution:  $R \approx 3 \times 10^{15}(t)^{0.9}$  cm,  $E_{K,R} \approx 10^{48}(t)^{0.8}$  erg and  $B \approx 2.4(t)^{-1}$  G, where  $B$  is the magnetic field strength (here  $t$  is in units of 5 d). The implied density profile is  $\rho(r) \propto r^{-2}$ , as expected for the wind from a massive star. **b**, Broadband radio spectra. The spectral peak of the radio synchrotron emission cascades to lower frequencies over the course of our follow-up observations with  $\nu_p \propto t^{-1}$ . The low frequency turn-over is consistent with expectations for synchrotron self-absorption (grey lines). **c**, Radio image from a VLBA observation on February 8 UT. The colour scale goes from  $-0.2$  mJy per beam (black) to  $1.4$  mJy per beam (white). We place an upper limit on the angular size of the ejecta of  $1.2$  mas ( $3\sigma$ ), corresponding to a physical radius of  $\lesssim 2.4 \times 10^{17}$  cm. This limit is a factor of 16 times larger than, and therefore consistent with, the radius derived from a massive star model. However, it places a limit of  $(\gamma\beta) \lesssim 3$  on the expansion velocity.

about two years. Along with the XRT field of view (24 arcmin on a side), the number density of  $L_*$  galaxies ( $\phi \approx 0.05 L_* \text{ Mpc}^{-3}$ ), and the detectability limit of XRT for events like XRO 080109 ( $d \lesssim 200$  Mpc), we infer an XRO rate of  $\gtrsim 10^{-3} L_*^{-1} \text{ yr}^{-1}$  (95% confidence level, Fig. 5); here  $L_*$  is the characteristic luminosity of galaxies<sup>37</sup>. This rate is at least an order of magnitude larger than for GRBs<sup>38,39</sup>. On the other hand, with a core-collapse supernova rate<sup>40</sup> of  $10^{-2} L_* \text{ yr}^{-1}$ , the probability of detecting at least one XRO if all such supernovae produce an outburst is about 50%.

We find a similar agreement with the supernova rate using the sensitivity of the BAT. The estimated<sup>39</sup> peak photon flux of the outburst is  $0.03 \text{ cm}^{-2} \text{ s}^{-1}$  (1–1,000 keV), which for a  $10^2$  s image trigger<sup>41</sup> is detectable to about 20 Mpc. The BAT on-sky monitoring time of 3 yr and the 2 sr field of view thus yield an upper limit on the XRO rate of  $\lesssim 10^5 \text{ Gpc}^{-3} \text{ yr}^{-1}$ , consistent with the core-collapse supernova rate<sup>42</sup> of  $6 \times 10^4 \text{ Gpc}^{-3} \text{ yr}^{-1}$ .



**Figure 5 | Volumetric rate of X-ray outbursts similar to XRO 080109.** We use all XRT observations longer than 300 s along with the field of view (24 arcmin on a side), the number density of  $L_*$  galaxies ( $\phi \approx 0.05 L_* \text{ Mpc}^{-3}$ ), and the detectability limit of XRT for events like XRO 080109 ( $d \lesssim 200$  Mpc). The curves indicate the rate ( $L_*^{-1} \text{ Mpc}^{-3} \text{ yr}^{-1}$ ) inferred from one detection in a total of about 2 yr of effective on-sky XRT observations as a function of the distance to which XROs can be detected. Also shown are the rates<sup>40</sup> of core-collapse supernovae (CC; solid horizontal line) and type Ibc supernovae (dashed horizontal line) as determined from optical supernova searches. The rate of events like XRO 080109 is consistent with the core-collapse rate at the 50% probability level.

Finally, we note that neither the late X-ray emission nor the radio emission show evidence for a rising component that could be attributed<sup>36</sup> to an off-axis GRB jet spreading into our line of sight. This conclusion is also supported by the unresolved size of the radio supernova from VLBA observations at  $t \approx 1$  month,  $R \lesssim 2.4 \times 10^{17}$  cm ( $3\sigma$ ), which constrains the outflow velocity to be  $\gamma\beta \lesssim 3$ .

### The rate of XROs

To estimate the rate of XROs, we find that the on-sky effective monitoring time of the XRT from the launch of Swift through to the end of January 2008, including only those exposures longer than 300 s, is

Finally, we note that NGC 2770 hosted an unusually high rate of three type Ibc supernovae in the past 10 yr. However, the galaxy has a typical luminosity ( $0.3 L_*$ ) and a total star formation rate of only  $0.5\text{--}1 M_\odot \text{ yr}^{-1}$  (see Supplementary Information), two orders of magnitude lower than the extreme starburst galaxy Arp 220, which has<sup>43</sup> a supernova rate of  $4 \pm 2 \text{ yr}^{-1}$ . The elevated supernova rate in NGC 2770, with a chance probability of  $\sim 10^{-4}$ , may simply be a statistical fluctuation, given the sample of  $\sim 4 \times 10^3$  known supernova host galaxies. Alternatively, it may point to a recent episode of increased star formation activity, perhaps triggered by interaction with the companion galaxy NGC 2770B at a separation<sup>44</sup> of only 22 kpc.

### Implications for supernova progenitors

Our observations probe the explosion ejecta over a wide range in velocity,  $\sim 10,000\text{--}210,000 \text{ km s}^{-1}$ . Taken together, the material giving rise to the X-ray outburst, the radio emission, and the optical light traces an ejecta profile of  $E_K \propto (\gamma\beta)^{-4}$  up to trans-relativistic velocities. This profile is in good agreement with theoretical expectations<sup>24</sup> for a standard hydrodynamic spherical explosion of a compact star, but much steeper<sup>39</sup> than for relativistic GRB-associated supernovae.

On the other hand, we note the similarity between the shock break-out properties of the He-rich SN 2008D and the He-poor GRB-associated SN 2006aj, both suggestive of a dense stellar wind around a compact Wolf-Rayet progenitor. In the context of type Ibc supernovae and GRB progenitors, this provides evidence for continuity (and probably a single progenitor system) between He-rich and He-poor explosions, perhaps including GRBs.

Looking forward, our inference that every core-collapse supernova is marked by an XRO places the discovery and study of supernovae on the threshold of a major change. An all-sky X-ray satellite with a sensitivity similar to that of the Swift/XRT would detect and localize several hundred core-collapse supernovae per year, even if they are obscured by dust, at the time of explosion. As we have shown here, this would enable a clear mapping between the properties of the progenitors and those of the supernovae. Most important, however, X-ray outbursts will provide an unprecedented positional and temporal trigger for neutrino and gravitational wave detectors (such as IceCube and Advanced LIGO), which may ultimately hold the key to unlocking the mystery of the supernova explosion mechanism, and perhaps the identity of the compact remnants.

Received 11 February; accepted 4 April 2008.

1. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24**, 205–253 (1986).
2. Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).
3. Arnett, W. D., Bahcall, J. N., Kirshner, R. P. & Woosley, S. E. Supernova 1987A. *Annu. Rev. Astron. Astrophys.* **27**, 629–700 (1989).
4. Colgate, S. A. Early gamma rays from supernovae. *Astrophys. J.* **187**, 333–336 (1974).
5. Klein, R. I. & Chevalier, R. A. X-ray bursts from Type II supernovae. *Astrophys. J.* **223**, L109–L112 (1978).
6. Waxman, E., Mészáros, P. & Campana, S. GRB 060218: A relativistic supernova shock breakout. *Astrophys. J.* **667**, 351–357 (2007).
7. Cappa, C., Goss, W. M. & van der Hucht, K. A. A Very Large Array 3.6 centimeter continuum survey of galactic Wolf-Rayet stars. *Astrophys. J.* **127**, 2885–2897 (2004).
8. Woosley, S. E., Hegar, A. & Weaver, T. A. The evolution and explosion of massive stars. *Astrophys. J.* **74**, 1015–1071 (2002).
9. MacFadyen, A. I., Woosley, S. E. & Hegar, A. Supernovae, jets, and collapsars. *Astrophys. J.* **550**, 410–425 (2001).
10. Soderberg, A. M. et al. The sub-energetic  $\gamma$ -ray burst GRB 031203 as a cosmic analogue to the nearby GRB 980425. *Nature* **430**, 648–650 (2004).
11. Page, K. L. et al. Observations of an X-ray transient in NGC 2770. *GCN Rep.* **110** (2008).
12. Deng, J. & Zhu, Y. Bright X-ray transient (a XRF?) in NGC 2770 – a SN optical counterpart? *GCN Circ.* **7160** (2008).
13. Valenti, S., Turatto, M., Navasardyan, H., Benetti, S. & Cappellaro, E. Early OT detection of XRF in NGC 2770 in asiago frames. *GCN Circ.* **7163** (2008).
14. Malesani, D. et al. Transient in NGC 2770: Spectroscopic evidence for a SN. *GCN Circ.* **7169** (2008).
15. Li, W. & Filippenko, A. V. Supernova 2008D in NGC 2770. *Cent. Bur. Electron. Telegr.* **1202** (2008).

16. Modjaz, M. et al. Supernova 2008D in NGC 2770. *Cent. Bur. Electron. Telegr.* **1222** (2008).
17. Pian, E. et al. An optical supernova associated with the X-ray flash XRF 060218. *Nature* **442**, 1011–1013 (2006).
18. Campana, S. et al. The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* **442**, 1008–1010 (2006).
19. Soderberg, A. M. et al. An HST study of the supernovae accompanying GRB 040924 and GRB 041006. *Astrophys. J.* **636**, 391–399 (2006).
20. Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497**, 17–20 (1998).
21. Waxman, E. & Loeb, A. TeV neutrinos and GeV photons from shock breakout in supernovae. *Phys. Rev. Lett.* **87**, 071101 (2001).
22. Wang, X.-Y., Li, Z., Waxman, E. & Mészáros, P. Nonthermal gamma-ray/X-ray flashes from shock breakout in gamma-ray burst-associated supernovae. *Astrophys. J.* **664**, 1026–1032 (2007).
23. Moffat, A. F. J., Drissen, L. & Robert, C. in *Physics of Luminous Blue Variables* (eds Davidson, K., Moffat, A. F. J. & Lamers, H. J. G. L. M.) 229–237 (IAU Colloq. 113, Kluwer Academic, Dordrecht, 1989).
24. Matzner, C. D. & McKee, C. F. The expulsion of stellar envelopes in core-collapse supernovae. *Astrophys. J.* **510**, 379–403 (1999).
25. Valenti, S. et al. The broad-lined Type Ic supernova 2003jd. *Mon. Not. R. Astron. Soc.* **383**, 1485–1500 (2008).
26. Arnett, W. D. Type I supernovae. I – Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).
27. Soderberg, A. M. et al. The radio and X-ray-luminous Type Ibc supernova 2003L. *Astrophys. J.* **621**, 908–920 (2005).
28. Berger, E., Kulkarni, S. R. & Frail, D. A. A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J.* **590**, 379–385 (2003).
29. Frail, D. A., Kulkarni, S. R., Berger, E. & Wieringa, M. H. A complete catalog of radio afterglows: The first five years. *Astron. J.* **125**, 2299–2306 (2003).
30. Berger, E., Kulkarni, S. R., Frail, D. A. & Soderberg, A. M. A radio survey of Type Ib and Ic supernovae: Searching for engine-driven supernovae. *Astrophys. J.* **599**, 408–418 (2003).
31. Kouveliotou, C. et al. Chandra observations of the X-ray environs of SN 1998bw/GRB 980425. *Astrophys. J.* **608**, 872–882 (2004).
32. Chevalier, R. A. Self-similar solutions for the interaction of stellar ejecta with an external medium. *Astrophys. J.* **258**, 790–797 (1982).
33. Readhead, A. C. S. Equipartition brightness temperature and the inverse Compton catastrophe. *Astrophys. J.* **426**, 51–59 (1994).
34. Kulkarni, S. R. et al. Radio emission from the unusual supernova 1998bw and its association with the  $\gamma$ -ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
35. Waxman, E. The nature of GRB 980425 and the search for off-axis gamma-ray burst signatures in nearby Type Ib/c supernova emission. *Astrophys. J.* **602**, 886–891 (2004).
36. Soderberg, A. M., Nakar, E., Berger, E. & Kulkarni, S. R. Late-time radio observations of 68 Type Ibc supernovae: Strong constraints on off-axis gamma-ray bursts. *Astrophys. J.* **638**, 930–937 (2006).
37. Blanton, M. R. et al. The galaxy luminosity function and luminosity density at redshift  $z = 0.1$ . *Astrophys. J.* **592**, 819–838 (2003).
38. Schmidt, M. Luminosity function of gamma-ray bursts derived without benefit of redshifts. *Astrophys. J.* **552**, 36–41 (2001).
39. Soderberg, A. M. et al. Relativistic ejecta from X-ray flash XRF 060218 and the rate of cosmic explosions. *Nature* **442**, 1014–1017 (2006).
40. Cappellaro, E., Evans, R. & Turatto, M. A new determination of supernova rates and a comparison with indicators for galactic star formation. **351**, 459–466 (1999).
41. Band, D. L. Postlaunch analysis of Swift's gamma-ray burst detection sensitivity. *Astrophys. J.* **644**, 378–384 (2006).
42. Dahlen, T. et al. High-redshift supernova rates. *Astrophys. J.* **613**, 189–199 (2004).
43. Lonsdale, C. J., Diamond, P. J., Thrall, H., Smith, H. E. & Lonsdale, C. J. VLBI images of 49 radio supernovae in Arp 220. *Astrophys. J.* **647**, 185–193 (2006).
44. Fynbo, J. P. U., Malesani, D., Auguestein, T. & Niemi, S.-M. NGC 2770B has the same redshift as NGC 2770. *GCN Circ.* **7186** (2008).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This Article is based in part on observations obtained at the Gemini Observatory through the Director's Discretionary Time. Gemini is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the NSF (US), the STFC (UK), the NRC (Canada), CONICYT (Chile), the ARC (Australia), CNPq (Brazil) and SECYT (Argentina). The VLA is operated by NRAO, a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. The Observatory was made possible by the financial support of the W. M. Keck Foundation. A.M.S. acknowledges support by NASA through a Hubble Fellowship.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.M.S. ([alicia@astro.princeton.edu](mailto:alicia@astro.princeton.edu)).



# Mechanism of shape determination in motile cells

Kinneret Keren<sup>1,3\*</sup>, Zachary Pincus<sup>1,4\*</sup>, Greg M. Allen<sup>1</sup>, Erin L. Barnhart<sup>1</sup>, Gerard Marriott<sup>5</sup>, Alex Mogilner<sup>6</sup> & Julie A. Theriot<sup>1,2</sup>

**The shape of motile cells is determined by many dynamic processes spanning several orders of magnitude in space and time, from local polymerization of actin monomers at subsecond timescales to global, cell-scale geometry that may persist for hours. Understanding the mechanism of shape determination in cells has proved to be extremely challenging due to the numerous components involved and the complexity of their interactions. Here we harness the natural phenotypic variability in a large population of motile epithelial keratocytes from fish (*Hypsophrys nicaraguensis*) to reveal mechanisms of shape determination. We find that the cells inhabit a low-dimensional, highly correlated spectrum of possible functional states. We further show that a model of actin network treadmilling in an inextensible membrane bag can quantitatively recapitulate this spectrum and predict both cell shape and speed. Our model provides a simple biochemical and biophysical basis for the observed morphology and behaviour of motile cells.**

Cell shape emerges from the interaction of many constituent elements—notably, the cytoskeleton, the cell membrane and cell–substrate adhesions—that have been studied in great detail at the molecular level<sup>1–3</sup>; however, the mechanism by which global morphology is generated and maintained at the cellular scale is not understood. Many studies have characterized the morphological effects of perturbing various cytoskeletal and other cellular components (for example, ref. 4); yet, there have been no comprehensive efforts to try to understand cell shape from first principles. Here we address this issue in the context of motile epithelial keratocytes derived from fish skin. Fish keratocytes are among the fastest moving animal cells, and their motility machinery is characterized by extremely rapid molecular dynamics and turnover<sup>5–8</sup>. At the same time, keratocytes are able to maintain nearly constant speed and direction during movement over many cell lengths. Their shapes, consisting of a bulbous cell body at the rear attached to a broad, thin lamellipodium at the front and sides, are simple, stereotyped and notoriously temporally persistent<sup>9,10</sup>. The molecular dynamism of these cells, combined with the persistence of their global shape and behaviour, make them an ideal model system for investigating the mechanisms of cell shape determination.

The relative simplicity of keratocytes has inspired extensive experimental and theoretical investigations into this cell type<sup>5–17</sup>, considerably advancing the understanding of cell motility. A notable example is the graded radial extension (GRE) model<sup>12</sup>, which was an early attempt to link the mechanism of motility at the molecular level with overall cell geometry. The GRE model proposed that local cell extension (either protrusion or retraction) occurs perpendicular to the cell edge, and that the magnitude of this extension is graded from a maximum near the cell midline to a minimum towards the sides. Although this phenomenological model has been shown experimentally to describe keratocyte motion, it does not consider what generates the graded extension rates, neither does it explain what determines the cellular geometry in the first place. Thus, even

for these simple cells, it has remained unclear how the biochemical and biophysical molecular dynamics underlying motility give rise to large-scale cell geometry. In this work we address this question by exploiting the natural phenotypic variability in keratocytes to measure the relations among cell geometry, actin distribution and motility. On the basis of quantitative observations of a large number of cells, we have developed a model that relates overall cell geometry to the dynamics of actin network treadmilling and the forces imposed on this network by the cell membrane. This model is able to quantitatively explain the main features of keratocyte shapes and to predict the relationship between cell geometry and speed.

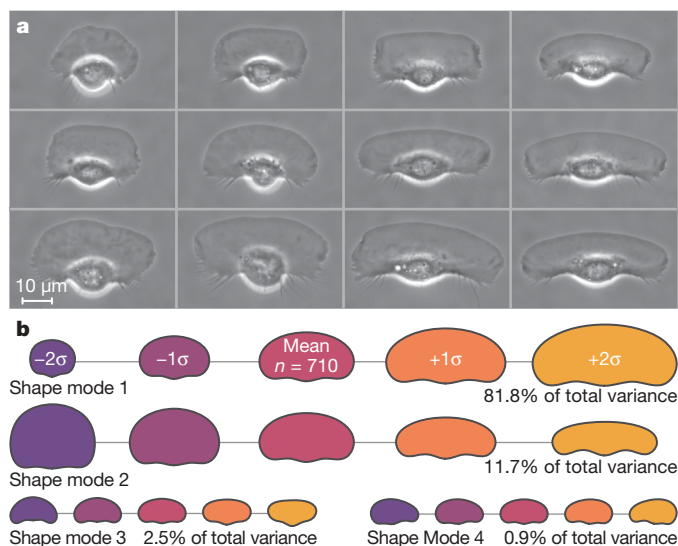
## Low-dimensional keratocyte shape space

Individual keratocytes assume a variety of cell shapes (Fig. 1a). A quantitative characterization<sup>18,19</sup> of a large population of live keratocytes revealed that keratocyte shapes are well described with just four orthogonal modes of shape variability (Fig. 1b), which together account for ~97% of the total variation in shape. Roughly, these modes can be characterized as measures of: the projected cell area (mode 1); whether the cell has a rounded ‘D’ shape or an elongated ‘canoe’ shape (mode 2)<sup>11</sup>; the angle of the rear of the lamellipodium with respect to the cell body (mode 3); and the left–right asymmetry of the side lobes (mode 4). These shape modes provide a meaningful and concise quantitative description of keratocyte morphology using very few parameters. Specifically, over 93% of the cell-to-cell shape variation can be captured by recording only two parameters per cell: the cell’s position along shape modes 1 and 2, or, essentially equivalently, its projected area and aspect ratio. Two additional parameters are required to describe the detailed shape of the rear of the cell (shape modes 3 and 4). The existence of only a few meaningful modes implies that the phase space in which keratocytes reside is a relatively small subregion of the space of all possible shapes.

To investigate further the role of various molecular processes in determining cell shape, we targeted specific components of the

<sup>1</sup>Department of Biochemistry, and <sup>2</sup>Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Department of Physics, Technion–Israel Institute of Technology, Haifa 32000, Israel. <sup>4</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA. <sup>5</sup>Department of Physiology, University of Wisconsin at Madison, Madison, Wisconsin 53706, USA. <sup>6</sup>Department of Neurobiology, Physiology and Behavior and Department of Mathematics, University of California, Davis, California 95616, USA.

\*These authors contributed equally to this work.



**Figure 1 | Keratocyte shapes are described by four primary shape modes.** **a**, Phase-contrast images of different live keratocytes illustrate the natural shape variation in the population. **b**, The first four principal modes of keratocyte shape variation, as determined by principal components analysis of 710 aligned outlines of live keratocytes, are shown. These modes—cell area (shape mode 1), ‘D’ versus ‘canoe’ shape (shape mode 2), cell-body position (shape mode 3), and left–right asymmetry (shape mode 4)—are highly reproducible; subsequent modes seem to be noise. For each mode, the mean cell shape is shown alongside reconstructions of shapes one and two standard deviations away from the mean in each direction along the given mode. The variation accounted for by each mode is indicated. (Modes one and two are scaled as in **a**; modes three and four are 50% smaller.)

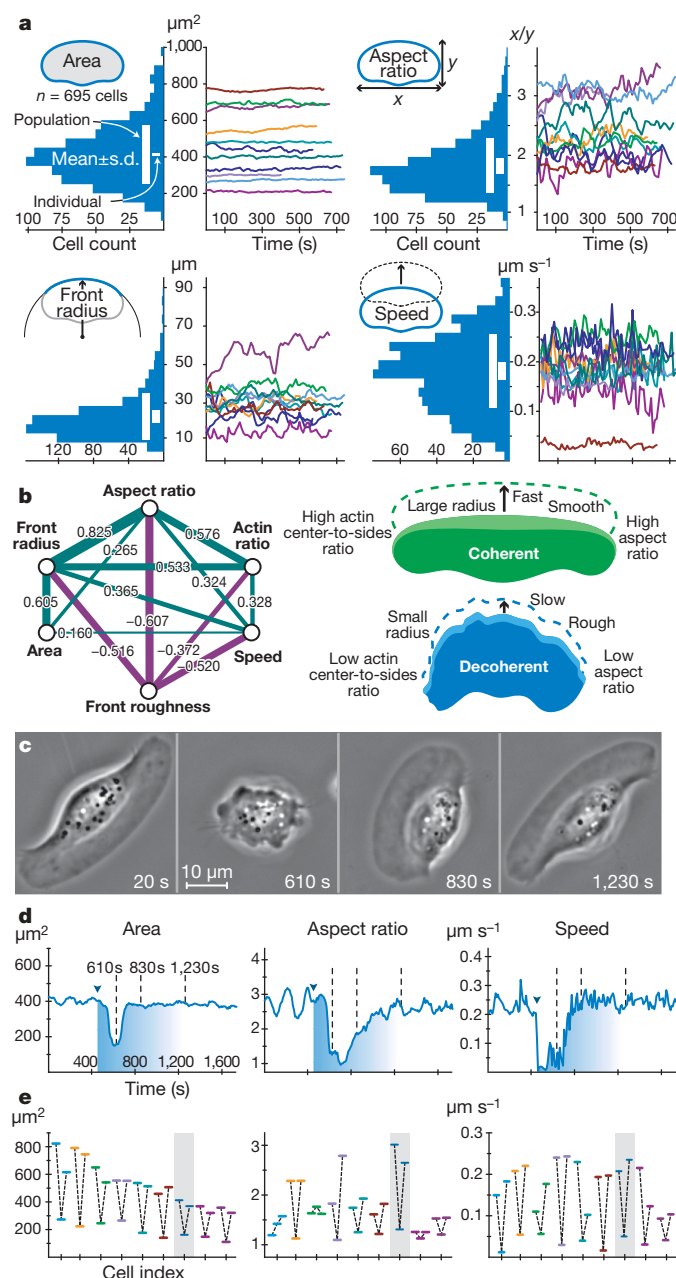
cytoskeleton in live cells with pharmacological agents that affect actin dynamics or myosin activity. The different treatments elicited statistically significant morphological changes (Supplementary Fig. 1), but their extent was rather small. In particular, the natural shape variation in the population (Fig. 1) was substantially larger than the shifts induced by any of the perturbations (Supplementary Fig. 1). Furthermore, whereas the shape of an individual cell can be significantly affected by such perturbations<sup>11</sup>, the phase space of cell

**Figure 2 | Quantitative and correlative analysis of keratocyte morphology and speed.** **a**, The distributions of measures across a population of live keratocytes (left panels) are contrasted with values through time for 11 individual cells (right). Within each histogram, the population mean  $\pm$  one standard deviation is shown by the left vertical bar, whereas the population mean  $\pm$  the average standard deviation exhibited by individual cells over 5 min is shown by the right bar. **b**, Significant pair-wise correlations ( $P < 0.05$ ; bootstrap confidence intervals) within a population of keratocytes are diagrammed (left panel). Two additional measures are included: front roughness, which measures the local irregularity of the leading edge, and actin ratio, which represents the peakedness of the actin distribution along the leading edge. The correlations indicate that, apart from size differences, cells lie along a single phenotypic continuum (right panel), from ‘decoherent’ to ‘coherent’. Decoherent cells move slowly and assume rounded shapes with low aspect ratios and high lamellipodial curvatures. The actin network is less ordered, with ragged leading edges and low actin ratios. Coherent cells move faster and have lower lamellipodial curvatures. The actin network is highly ordered with smooth leading edges and high actin ratios. **c**, Phase-contrast images depict a cell transiently treated with DMSO (Supplementary Movie 1), which caused a reversible inhibition of motility and loss of the lamellipodium. Images shown correspond to before (20 s), during (610 s) and two time points after (830 s and 1,230 s) the perturbation. **d**, Time traces of area, aspect ratio and speed for the cell in **c** show that shape and speed are regained post perturbation. Dashed lines show time points from **c**; arrowheads indicate the time of perturbation. **e**, Area, aspect ratio and speed of nine cells are shown as averages obtained from one-minute windows before, during and after DMSO treatment (shown sequentially from left to right for each cell). The cell shown in **c** and **d** is highlighted.

shapes under the perturbations tested was nearly identical to that spanned by the population of unperturbed cells (Supplementary Fig. 1). This led us to focus on the phenotypic variability in unperturbed populations, which, as described, provided significant insight into the underlying mechanisms of shape determination.

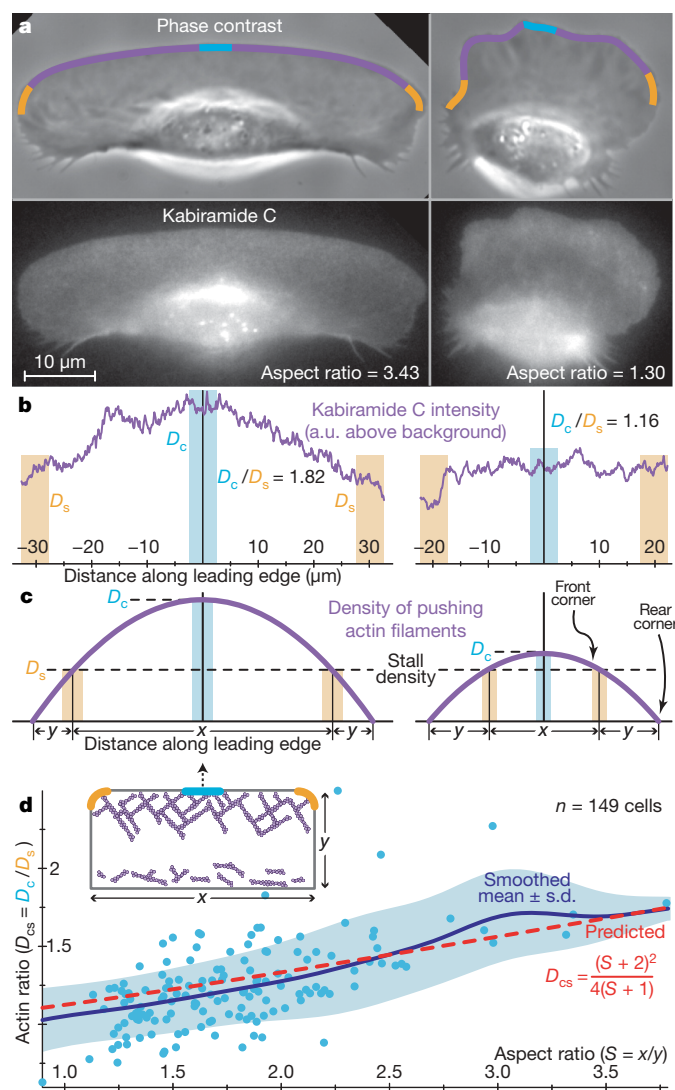
### Cell shape is dynamically determined

The natural phenotypic variability described presents a spectrum of possible functional states of the system. To better characterize these states, we measured cell speed, area, aspect ratio and other morphological features in a large number of live cells (Fig. 2a) and correlated these traits across the population (Fig. 2b; see also the Supplementary Information). To relate these measures to cellular actin dynamics, we concurrently examined the distribution of actin filaments along the leading edge. To visualize actin filaments in live cells, we used low levels of tetramethylrhodamine (TMR)-derivatized kabiramide C, which at low concentrations binds as a complex with G-actin to free barbed ends of actin filaments<sup>20,21</sup>, so that along the leading edge the measured fluorescence intensity is proportional to the local density of filaments.





The phenotypic variability in our test population is depicted in the histograms shown in Fig. 2a. We further characterized this variability by following several individual cells over time. Particularly notable



**Figure 3 | A quantitative model explains the main features of keratocyte shapes.** **a**, Phase-contrast (top) and fluorescence (bottom) images are shown for two live keratocytes stained with TMR-derivatized kabiramide C. The fluorescence intensity reflects the current and past distribution of filament ends, in addition to diffuse background signal from unincorporated probe<sup>20</sup>. Along the leading edge, the fluorescence intensity is proportional to the local density of actin filaments (see Supplementary Information; 1-μm-wide strips along the leading edge are shown superimposed on the phase-contrast images, with centre and side regions highlighted). **b**, The average (background-corrected) fluorescence intensity along the strips shown in **a** is plotted. The cell on the left has a peaked distribution of actin filaments, whereas the actin distribution in the cell on the right is flatter. The ratio of the actin density at the centre ( $D_c$ ) and sides ( $D_s$ ; averaged over both sides) of the strip, denoted as  $D_{cs}$ , serves as a robust measure of the peakedness of the distribution. **c**, The density distribution of pushing actin filaments along the leading edge is approximated as a parabola, with a maximum at the centre. Cells with peaked filamentous actin distributions and, therefore, high  $D_{cs}$  values, have larger regions in which the actin filament density is above the 'stall' threshold, and thus have longer protruding front edges (of length  $x$ ) compared with the length of the stalled/retracting cell sides ( $y$ ), yielding higher aspect ratios ( $S = x/y$ ). **d**, The ratio between actin density at the centre and at the sides,  $D_{cs}$ , is plotted as a function of cell aspect ratio,  $S$ . Each data point represents an individual cell. Our model provides a parameter-free prediction of this relationship (red line), which captures the mean trend in the data, plotted as a gaussian-weighted moving average ( $\sigma = 0.25$ ; blue line)  $\pm$  one standard deviation (blue region). Inset: the model of cell shape is illustrated schematically.

was the observation that the projected cell area, although quite variable across the population, was essentially constant for a given cell (Fig. 2a). This suggests that the area, probably determined by the total amount of available plasma membrane or by tight regulation of the membrane surface area, is intrinsic to each cell and constant through time. Individual cells showed larger variability in other measures such as speed and aspect ratio; nevertheless, in every case, individual variability remained smaller than that of the population as a whole (Fig. 2a). The measured properties correlate well across the data set (Fig. 2b and Supplementary Fig. 2), producing a phenotypic continuum that we have described previously<sup>11</sup>: from rough, slow and rounded 'decoherent' cells, to smooth, fast and wide 'coherent' cells that exhibit a more pronounced peak in actin filament density at the centre.

To examine the role that the particular history of a given cell has in determining cell morphology, we confronted keratocytes with an acute perturbation—transient treatment with high concentrations of dimethylsulphoxide (DMSO)—which resulted in temporary lamellipodial loss and cell rounding<sup>22</sup>. We found that cells were able to resume movement (albeit in an arbitrary direction with respect to their orientation before DMSO treatment) and return to their original morphology and speed within minutes (Fig. 2c–e), comparable to the characteristic timescales of the underlying molecular processes such as actin assembly and disassembly and adhesion formation<sup>5–8,23</sup>. This rapid recovery of pre-perturbation properties suggests that the observed, persistent behaviour of keratocytes is a manifestation of a dynamic system at steady state. Taken together, our results imply that cell shape and speed are determined by a history-independent self-organizing mechanism, characterized by a small number of cellular parameters that stay essentially constant over time (such as available quantities of membrane or cytoskeletal components), independent of the precise initial localization of the components of the motility machinery.

### Actin/membrane model explains cell shape

We set out to develop a quantitative physical model of cell shape and movement that could explain this observed spectrum of keratocyte behaviour. Specifically, we sought to describe mechanistically the shape variability captured in the first two principal modes of keratocyte shape (Fig. 1b; comprising over 93% of the total shape variation), setting aside the detailed shape of the cell rear. Two observations—first, that cell area is constant (Fig. 2a), and second, that the density of filamentous actin along the leading edge is graded (Fig. 3a,b)—are central to our proposed mechanism of cell shape regulation. In addition, this mechanism is predicated on the basis of previous observations that the lamellipodial actin network undergoes treadmilling, with net assembly at the leading edge and net disassembly towards the rear<sup>8,24,25</sup>.

We hypothesize that actin polymerization pushes the cell membrane from within, generating membrane tension<sup>26</sup>. The cell membrane, which has been observed to remain nearly stationary in the cell frame of reference in keratocytes<sup>12,14</sup>, is fluid and bends easily but is nevertheless inextensible (that is, it can be deformed but not stretched)<sup>27</sup>. Forces on the membrane at any point equilibrate within milliseconds<sup>26</sup> (see Supplementary Information) so that, on the time-scales relevant for motility, membrane tension is spatially homogeneous at all points along the cell boundary. At the leading edge, membrane tension imposes an opposing force on growing actin filaments that is constant per unit edge length, so that the force per filament is inversely proportional to the local filament density. At the centre of the leading edge, where filament density is high (Fig. 3a–c), the membrane resistance per filament is small, allowing filaments to grow rapidly and generate protrusion. As filament density gradually decreases towards the cell sides, the forces per filament caused by membrane tension increase until polymerization is stalled at the far sides of the cell, which therefore neither protrude nor retract. At the rear of the cell, where the actin network disassembles,

membrane tension, assisted by myosin contraction, crushes the weakened network and moves actin debris forward, thereby retracting the cell rear (Fig. 3d, inset). Membrane tension, which is spatially constant, thus induces a direct coupling between molecular processes occurring at distant regions of the cell and contributes to the global coordination of those processes. The Supplementary Information discusses alternative hypotheses regarding cell shape determination that are inconsistent with our measurements (Supplementary Fig. 3).

This qualitative model can be mathematically specified and quantitatively compared to our data set as follows (see Supplementary Table 1 for a list of model assumptions, and Supplementary Information for further details). As discussed previously (Fig. 1), keratocyte shapes can largely be described by two parameters: shape modes 1 and 2, which essentially correspond to cell area ( $A$ ) and aspect ratio ( $S$ ), respectively. Thus, for simplicity, we begin by approximating cells as rectangles with width  $x$  and length  $y$  ( $A = xy$ ,  $S = x/y$ , and the total leading edge length (front and sides) is  $L = x + 2y = \sqrt{AS} + 2\sqrt{A/S}$ ). The observed steady-state centre-peaked distribution of actin filaments along the leading edge ( $D$ ) can be described as a parabola:  $D(l) = \frac{\beta}{L\gamma} \left( 1 - \left( \frac{l}{L/2} \right)^2 \right)$ , where  $l$  is

the arc distance along the leading edge ( $l = 0$  at the cell midline),  $\beta$  is the total number of nascent actin filaments that branch off from existing growing filaments per cell per second, and  $\gamma$  is the rate of capping of existing filaments (Fig. 3c; see Supplementary Information for derivation). We make the further assumption (described previously) that actin filament protrusion is mechanically stalled by the membrane tension  $T$  at the sides of the front of the lamellipodium ( $l = \pm x/2$ ). The force acting on each filament at the sides must therefore be approximately equal to the force required to stall a single actin filament<sup>28</sup>,  $f_{\text{stall}}$ , which has been measured<sup>29,30</sup>, so that:  $D_s = D(x/2) = \frac{\beta}{L\gamma} \left( 1 - \left( \frac{x}{L} \right)^2 \right) = \frac{T}{f_{\text{stall}}}$ . We find that the peak actin density  $D_c = D(0)$  fluctuates more than  $D_s$  across the population and in individual cells through time (Supplementary Fig. 4; Supplementary Information), suggesting that most of the shape variation observed correlates with differences in actin dynamics rather than changes in membrane tension.

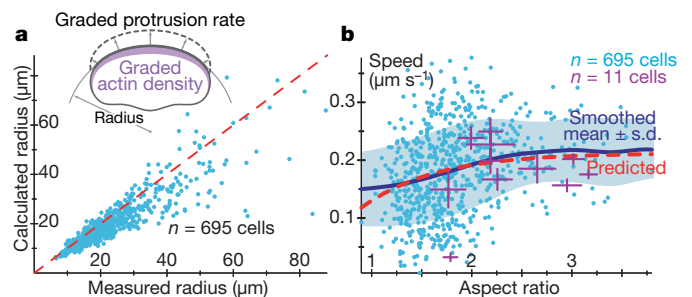
This simple model provides a direct link between the distribution of filamentous actin and overall cell morphology. From the previous equations, this link can be expressed as a relation between the ratio of actin filament density at the centre ( $l = 0$ ) versus the sides ( $l = \pm x/2$ ) of the leading edge, denoted  $D_{cs}$ , and the aspect ratio of the cell,  $S$ :  $D_{cs} = \frac{D_c}{D_s} = \left[ 1 - \left( \frac{x}{L} \right)^2 \right]^{-1} = \frac{(S+2)^2}{4(S+1)}$ . Thus, cells with relatively more actin filament density at the centre than the sides (high  $D_{cs}$ ) have higher aspect ratios, whereas cells with low  $D_{cs}$  ratios have aspect ratios closer to one. As shown in Fig. 3d, the correlation between  $D_{cs}$  and  $S$  in our measurements closely follows this model prediction, which, importantly, involves no free parameters. The model is further supported by perturbation experiments, in which, for example, increasing the capping rate  $\gamma$  (by treatment with cytochalasin D) led to the predicted decrease in cell aspect ratio (Supplementary Fig. 1; Supplementary Information). Remarkably, all the model parameters apart from area can be combined into a single parameter:  $z = \frac{T\gamma}{f_{\text{stall}}\beta}$ , which signifies the ratio of the membrane tension to the force needed to stall actin network growth at the centre of the leading edge. This key parameter can be expressed in multiple ways:  $z \equiv \frac{T\gamma}{f_{\text{stall}}\beta} = \frac{1}{L} \left( 1 - \left( \frac{x}{L} \right)^2 \right) = \frac{1}{L D_{cs}}$ ; that is, in terms of the membrane tension, filament stall force, and branching and capping rates; in terms of the measurable geometry of the cell alone; or in terms of the actin density ratio and cell geometry (see also Supplementary Fig. 5). Thus, this model describes the basic relation between actin network dynamics at the molecular level and overall actin network structure and shape at the cellular scale using only two biologically relevant parameters:  $z$  and  $A$ .

### Shape, speed and lamellipodial radius

To describe cell shape with more accuracy and to relate cell speed to morphology, we must consider the relationship between the growth rate of actin filaments and the magnitude of force resisting their growth. This so-called force-velocity relationship can be used to determine the protrusion rate at the leading edge, and thus cell speed, from the forces exerted by the membrane against the growing lamellipodial actin network. Because membrane tension is the same everywhere along the leading edge, although the filamentous actin density is peaked at the centre of the leading edge, the resistive force per filament increases with distance from the centre. As a result, local protrusion rates decrease smoothly from the centre towards the sides of the leading edge (where, as above, protrusion is stalled). Assuming that protrusion is locally perpendicular to the cell boundary, this implies that the sides of the leading edge lag behind the centre, causing the leading edge to become curved as observed (Fig. 1a; such a relation between geometry and spatially variable protrusion rates was first described in the GRE model<sup>12</sup>). Thus, keratocytes can be more accurately described as slightly bent rectangles, characterized by the radius of curvature of their leading edge,  $R$ , and their overall rate of movement (Fig. 4), in addition to their width and length.

Given a particular force-velocity relation, both cell speed and lamellipodial radius can be expressed, in the context of this model, solely in terms of the parameters  $A$  and  $z$ . Thus, speed and radius are predicted to vary with cell area and aspect ratio, providing further tests of the model. The exact form of the force-velocity relation for the lamellipodial actin network is unknown. Measurements in branched actin networks, both in motile keratocytes<sup>16</sup> and assembled in cytoplasmic extracts<sup>31</sup>, yielded force-velocity relations that were concave down: that is, the protrusion rate was insensitive to force at weak loads (relative to the stall force), whereas at greater loads the speed decreased markedly. Regardless of its precise functional dependence, as long as the force-velocity relation entails such a monotonic concave-down decrease in protrusion velocity with increasing membrane tension, the predicted trends in cell speed and lamellipodium radius correlate well with our experimental observations (Supplementary Fig. 6). We find good quantitative agreement between the model and our observations using a force-velocity relation given by  $V = V_0 \left( 1 - \left( \frac{f}{f_{\text{stall}}} \right)^w \right)$ , where  $w = 8$  (Fig. 4).

By combining this force-velocity relation with the geometric formulae of the GRE model, we obtain  $R \approx \frac{L}{8} \sqrt{(zL)^{-8} - 1}$  (see Supplementary



**Figure 4 | An extended model predicts lamellipodial curvature and the relationship between speed and morphology.** **a**, The radius of curvature of the leading edge calculated within the model as a function of  $A$  and  $S$ ,  $R_c = \frac{L}{8} \sqrt{(zL)^{-8} - 1}$ , with  $zL = \frac{4(S+1)}{(S+2)^2}$  and  $L = \sqrt{AS} + 2\sqrt{A/S}$ , is plotted against the measured radius of curvature ( $R_m$ , radius of best-fit circle of the front 40% of the cell). The red dashed line depicts  $R_c = R_m$ . **b**, Cell speed,  $V_{\text{cell}}$ , is shown as a function of cell aspect ratio,  $S$ . The model prediction  $V_{\text{cell}} = V_0 \left( 1 - \left( \frac{4(S+1)}{(S+2)^2} \right)^8 \right)$  (red line;  $V_0$  determined empirically) is compared to the trend plotted as a gaussian-weighted moving average ( $\sigma = 0.25$ ; blue line)  $\pm$  one standard deviation (blue region), from 695 individual cells (blue points). Purple crosses indicate the mean  $\pm$  one standard deviation in speed and aspect ratio over 5 min for 11 individual cells (shown in Fig. 2a).

Information), which predicts the radius of curvature of a cell's leading edge from its area and aspect ratio alone. Figure 4a demonstrates the close agreement between the measured and the calculated radii of curvature. At the centre of the leading edge,  $f = T/D_c$ ; therefore,  $V_{\text{cell}} = V_0 \left( 1 - \left( \frac{T}{f_{\text{sat}} D_c} \right)^8 \right) = V_0 (1 - (zL)^8) = V_0 \left( 1 - \left( \frac{4(S+1)}{(S+2)^2} \right)^8 \right)$ .

Thus, a cell's speed can be predicted from its aspect ratio, with more canoe-like cells expected to move faster. We find that the trend of the experimental data agrees with our predictions (Fig. 4b), and, in particular, shows the predicted saturation of speed with increasing aspect ratio. We expect cell-to-cell variation in some of the model parameters that determine cell speed such as the concentration of actin monomers and the fraction of pushing actin filaments, as well as in the rate of retrograde actin flow with respect to the substrate<sup>13,17</sup>. Without detailed per-cell measurements of these, we use constant values that reflect the population mean, allowing correct prediction of population trends, whereas some aspects of cell-to-cell variation remain unexplained.

## Discussion

We have used correlative approaches to map quantitatively the functional states of keratocyte motility from a large number of observations of morphology, speed and actin network structure in a population of cells. This data set provided the basis for and constraints on a quantitative model of cell shape that requires only two cell-dependent parameters; these parameters are measurable from cell geometry alone and are closely related to the two dimensions of a phase space that accounts for over 93% of all keratocyte shape variation. Although conceptually quite straightforward, our model describes connections between dynamic events spanning several orders of magnitude in space and time and is, to our knowledge, the first quantitative approach relating molecular mechanisms to cell geometry and movement. The model is able to explain specific properties of keratocyte shape and locomotion on the basis of a coupling of tension in the cell membrane to the dynamics of the treadmilling network of actin filaments. Overall, the picture is very simple: actin network treadmilling (characterized by the  $z$  parameter) drives from within the forward protrusion of an inextensible membrane bag (characterized in two dimensions by its total area). Such a scenario was suggested over a decade ago<sup>32</sup>, but prior to this work had never been tested. Furthermore, this basic mechanism seems to be sufficient to explain the persistent and coordinated movement of keratocytes without incorporating regulatory elements such as microtubules, morphogens or signalling molecules<sup>33</sup>, suggesting that, at least in keratocytes, these elements are dispensable or redundant.

The model highlights the important regulatory role of membrane tension in cell shape determination: actin assembly at the leading edge and disassembly at the cell rear are both modulated by forces imposed on the actin network by the membrane. Moreover, because membrane tension is constant along the cell boundary, it effectively couples processes (such as protrusion and retraction) that take place in spatially distinct regions of the cell. On the basis of our results, we estimate the membrane tension in motile keratocytes to be on the order of  $100 \text{ pN } \mu\text{m}^{-1}$  (see Supplementary Information), similar to the results of experiments that estimated membrane tension from the force on a tether pulled from the surface of motile fibroblasts<sup>34</sup>.

Our model does not specifically address adhesion or the detailed shape of the cell rear (captured in shape modes 3 and 4; Fig. 1b). Nevertheless, adhesive contacts to the substrate are obviously essential for the cell to be able to generate traction and to move forward. We assume implicitly that the lamellipodial actin network is attached to the substrate, which allows polymerization to translate into cellular protrusion. This assumption is consistent with experimental evidence indicating that the actin network in the keratocyte lamellipodium is nearly stationary with respect to the substrate<sup>8,13,17</sup>. The rear boundary of the cell is also implicit in our model, and is set by the position of the 'rear corners' of the lamellipodium: the locations at

which the density of actin filaments actively pushing against the cell membrane falls to zero. Thus, we do not address the possible contribution of myosin contraction in retracting the cell rear and disassembling the actin network<sup>7,26</sup> (see Supplementary Information).

Our results emphasize that careful quantitative analysis of natural cell-to-cell variation can provide powerful insight into the molecular mechanisms underlying complex cell behaviour. A rapidly moving keratocyte completely rebuilds its cytoskeleton and adhesive structures every few minutes, generating a cell shape that is both dynamically determined and highly robust. This dynamic stability suggests that shape emerges from the numerous molecular interactions as a steady-state solution, without any simple central organizing or book-keeping mechanism. In this work, we relied on several decades of detailed mechanistic studies on the molecular mechanisms involved to derive a physically realistic model for large-scale shape determination. This model is directly and quantitatively coupled to the molecular-scale dynamics and has surprising predictive power. As individual functional modules within cells are unveiled at the molecular level, understanding their large-scale integration is becoming an important challenge in cell biology. To this end, we propose that the biologically rich cell-to-cell variability present within all normal populations represents a fruitful but currently underused resource of mechanistic information regarding complex processes such as cell motility.

## METHODS SUMMARY

**Cell culture.** Keratocytes were isolated from the scales of the Central American cichlid *H. nicaraguensis* and were cultured as described previously<sup>11</sup>. TMR-derivatized kabiramide C was added to cells in culture medium for 5 min and subsequently washed<sup>20</sup>. DMSO treatment consisted of either application of 2–5  $\mu\text{l}$  DMSO directly onto cells or addition of 10% DMSO to the culture medium.

**Microscopy.** Cells were imaged in a live-cell chamber at room temperature ( $\sim 23^\circ\text{C}$ ) on a Nikon Diaphot300 microscope using a  $\times 60$  lens (numerical aperture, 1.4). To obtain velocity information, for each coverslip, 15–30 randomly chosen cells were imaged twice, 30 s apart. Time-lapse movies of individual cells were acquired at 10-s intervals.

**Shape analysis.** Cell morphology was measured from manually defined cell shapes, as described previously<sup>11,19</sup>. 'Shape modes' were produced by performing principal components analysis on the population of cell shapes after mutual alignment.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 14 December 2007; accepted 31 March 2008.

1. Carlier, M. F. & Pantaloni, D. Control of actin assembly dynamics in cell motility. *J. Biol. Chem.* **282**, 23005–23009 (2007).
2. Pollard, T. D., Blanchoin, L. & Mullins, R. D. Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576 (2000).
3. Zaidel-Bar, R., Cohen, M., Addadi, L. & Geiger, B. Hierarchical assembly of cell-matrix adhesion complexes. *Biochem. Soc. Trans.* **32**, 416–420 (2004).
4. Bakal, C., Aach, J., Church, G. & Perrimon, N. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* **316**, 1753–1756 (2007).
5. Anderson, K. I. & Cross, R. Contact dynamics during keratocyte motility. *Curr. Biol.* **10**, 253–260 (2000).
6. Lee, J. & Jacobson, K. The composition and dynamics of cell–substratum adhesions in locomoting fish keratocytes. *J. Cell Sci.* **110**, 2833–2844 (1997).
7. Svitkina, T. M., Verkhovsky, A. B., McQuade, K. M. & Borisy, G. G. Analysis of the actin–myosin II system in fish epidermal keratocytes: mechanism of cell body translocation. *J. Cell Biol.* **139**, 397–415 (1997).
8. Theriot, J. A. & Mitchison, T. J. Actin microfilament dynamics in locomoting cells. *Nature* **352**, 126–131 (1991).
9. Euteneuer, U. & Schliwa, M. Persistent, directional motility of cells and cytoplasmic fragments in the absence of microtubules. *Nature* **310**, 58–61 (1984).
10. Goodrich, H. B. Cell behavior in tissue cultures. *Biol. Bull.* **46**, 252–262 (1924).
11. Lacayo, C. I. et al. Emergence of large-scale cell morphology and movement from local actin filament growth dynamics. *PLoS Biol.* **5**, e233 (2007).
12. Lee, J., Ishihara, A., Theriot, J. A. & Jacobson, K. Principles of locomotion for simple-shaped cells. *Nature* **362**, 167–171 (1993).



13. Jurado, C., Haserick, J. R. & Lee, J. Slipping or gripping? Fluorescent speckle microscopy in fish keratocytes reveals two different mechanisms for generating a retrograde flow of actin. *Mol. Biol. Cell* **16**, 507–518 (2005).
14. Kucik, D. F., Elson, E. L. & Sheetz, M. P. Cell migration does not produce membrane flow. *J. Cell Biol.* **111**, 1617–1622 (1990).
15. Grimm, H. P., Verkhovsky, A. B., Mogilner, A. & Meister, J. J. Analysis of actin dynamics at the leading edge of crawling cells: implications for the shape of keratocyte lamellipodia. *Eur. Biophys. J.* **32**, 563–577 (2003).
16. Prass, M., Jacobson, K., Mogilner, A. & Radmacher, M. Direct measurement of the lamellipodial protrusive force in a migrating cell. *J. Cell Biol.* **174**, 767–772 (2006).
17. Vallotton, P. *et al.* Tracking retrograde flow in keratocytes: news from the front. *Mol. Biol. Cell* **16**, 1223–1231 (2005).
18. Coates, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. Active shape models — their training and application. *Comput. Vis. Image Underst.* **61**, 38–59 (1995).
19. Pincus, Z. & Theriot, J. A. Comparison of quantitative methods for cell-shape analysis. *J. Microsc.* **227**, 140–156 (2007).
20. Petchprayoon, C. *et al.* Fluorescent kabiramides: new probes to quantify actin *in vitro* and *in vivo*. *Bioconjug. Chem.* **16**, 1382–1389 (2005).
21. Tanaka, J. *et al.* Biomolecular mimicry in the actin cytoskeleton: mechanisms underlying the cytotoxicity of kabiramide C and related macrolides. *Proc. Natl Acad. Sci. USA* **100**, 13851–13856 (2003).
22. Sanger, J. W., Gwinn, J. & Sanger, J. M. Dissolution of cytoplasmic actin bundles and the induction of nuclear actin bundles by dimethyl sulfoxide. *J. Exp. Zool.* **213**, 227–230 (1980).
23. Watanabe, N. & Mitchison, T. J. Single-molecule speckle analysis of actin filament turnover in lamellipodia. *Science* **295**, 1083–1086 (2002).
24. Pollard, T. D. & Borisy, G. G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**, 453–465 (2003).
25. Wang, Y. L. Exchange of actin subunits at the leading edge of living fibroblasts: possible role of treadmilling. *J. Cell Biol.* **101**, 597–602 (1985).
26. Kozlov, M. M. & Mogilner, A. Model of polarization and bistability of cell fragments. *Biophys. J.* **93**, 3811–3819 (2007).
27. Sheetz, M. P., Sable, J. E. & Dobereiner, H. G. Continuous membrane–cytoskeleton adhesion requires continuous accommodation to lipid and cytoskeleton dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 417–434 (2006).
28. Schaus, T. E. & Borisy, G. Performance of a population of independent filaments in lamellipodial protrusion. *Biophys. J.* (in the press).
29. Footer, M. J., Kerssemakers, J. W., Theriot, J. A. & Dogterom, M. Direct measurement of force generation by actin filament polymerization using an optical trap. *Proc. Natl Acad. Sci. USA* **104**, 2181–2186 (2007).
30. Kovar, D. R. & Pollard, T. D. Insertional assembly of actin filament barbed ends in association with formins produces piconewton forces. *Proc. Natl Acad. Sci. USA* **101**, 14725–14730 (2004).
31. Parekh, S. H., Chaudhuri, O., Theriot, J. A. & Fletcher, D. A. Loading history determines the velocity of actin-network growth. *Nature Cell Biol.* **7**, 1219–1223 (2005).
32. Mitchison, T. J. & Cramer, L. P. Actin-based cell motility and cell locomotion. *Cell* **84**, 371–379 (1996).
33. Ridley, A. J. *et al.* Cell migration: integrating signals from front to back. *Science* **302**, 1704–1709 (2003).
34. Raucher, D. & Sheetz, M. P. Cell spreading and lamellipodial extension rate is regulated by membrane tension. *J. Cell Biol.* **148**, 127–136 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Lacayo, C. Wilson and M. Kozlov for discussion, and P. Yam, C. Lacayo, E. Braun and T. Pollard for comments on the manuscript. K.K. is a Damon Runyon Postdoctoral Fellow supported by the Damon Runyon Cancer Research Foundation, and a Horev Fellow supported by the Taub Foundations. A.M. is supported by the National Science Foundation grant number DMS-0315782 and the National Institutes of Health Cell Migration Consortium grant number NIGMS U54 GM64346. J.A.T. is supported by grants from the National Institutes of Health and the American Heart Association.

**Author Contributions** Z.P., K.K., E.L.B., G.M.A. and J.A.T. designed the experiments. K.K., G.M.A., E.L.B. and Z.P. performed the experiments. Z.P. together with K.K., A.M., G.M.A. and E.L.B. analysed the data. A.M. together with K.K., Z.P., E.L.B., G.M.A. and J.A.T. developed the model. G.M. provided the kabiramide C probe. Z.P., K.K., A.M. and J.A.T. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.A.T. ([theriot@stanford.edu](mailto:theriot@stanford.edu)).

## METHODS

**Cell culture.** Keratocyte sheets from one-day-old cultures were disaggregated by incubating in 85% PBS and 2.5 mM EGTA, pH 7.4, for 5 min, followed by incubation in normal media for an additional ~1–2 h. TMR-derivatized kabiramide C was added to cells in culture medium for 5 min and subsequently washed<sup>20</sup>. Pharmacological agents including, cytochalasin D (Sigma), latrunculin, jasplakinolide (both from Molecular Probes), blebbistatin (active enantiomer, Toronto Research Chemicals) or calyculin A (Upstate), were applied to cells in culture medium, and the cells were imaged 10–30 min afterwards.

**Microscopy.** Images were collected on a cooled back-thinned CCD camera (Princeton Instruments), with a  $\times 2$  optovar attached (1 pixel = 0.11  $\mu\text{m}$ ). The population data was acquired by imaging 15–30 randomly chosen cells per coverslip.

**Shape analysis.** Cell morphology was measured from cell shapes represented as polygonal outlines and mutually aligned, as described previously<sup>11,19</sup>. In brief, cell shapes were manually masked using the magnetic-lasso tool in Adobe Photoshop on the phase-contrast image and stored as binary images. Polygonal outlines were extracted from these masks and represented as two-dimensional parametric periodic uniform cubic B-splines, which were sampled at 200 evenly spaced points to generate the final polygons. These were then aligned across the population to ensure that all polygons were oriented similarly; to facilitate this, the centroid of the cell body—a landmark by which the front and rear of the cell can be automatically determined—was extracted from the fluorescent kabiramide C image or by manual marking. Simultaneously, the point ordering of each polygon was adjusted so that corresponding points were in similar spatial locations on the cell across the population. (See algorithms 1 and 2 in Supplementary Information for details.) Cell alignment was then manually verified. The ‘shape modes’ were produced by applying the principal components analysis to the population of cell shapes, represented as 400-dimensional vectors of packed ( $x$ ,  $y$ ) points, and scaled in terms of the standard deviation of the population of shapes along that principle component.

Measured cellular characteristics included: cell area; aspect ratio; lamellipodial radius; speed; front roughness; and actin ratio. Area was measured directly from the polygons with the standard formula. Aspect ratio was measured as the ratio of the width to the length of the cell’s bounding box after cells were mutually aligned as above. The roughness of the leading edge of each cell was measured by calculating the average absolute value of the local curvature at each point along the leading edge, corrected for effects due to cell size<sup>11</sup>. The overall curvature of the leading edge was calculated as the radius of the least-squares ‘geometric fit’ of a circle to the points corresponding to the leading edge (the forward 40% of the cell)<sup>35</sup>. The distribution of kabiramide C staining along the leading edge was calculated by averaging the intensity of background-corrected fluorescence images between the cell edge (as determined by the polygon) and 1  $\mu\text{m}$  inward from there. The centre intensity was defined as the average of this profile in a 5- $\mu\text{m}$ -wide window centred on the cell midline; side intensity was defined as the average in similar windows at the left and right sides of the cell. Cell speed for the live population data was extracted from the displacement of the cell centroid as determined from the manually drawn masks of the two images taken 30 s apart for each cell. Angular cell speed was extracted from the relative rotation angle required for alignment of the two cell shapes. For time-lapse movies of individual cells and DMSO-treated cells taken with a 10-s time interval, the centroid based measurements were noisy so we relied on a correlation-based technique<sup>36</sup>. The translation and rotation of a cell between a pair of consecutive time-lapse images were extracted as in ref. 36, with the modification that the masks used were based on the manually drawn cell masks and the centre of rotation was taken as the centroid of the mask in the first image. All measurements of individual cells (unstained, stained with kabiramide C, and perturbed, as well as a fixed-cell population) and on cells followed with time-lapse microscopy (stained with kabiramide C and perturbed with DMSO) are provided as Supplementary Tables.

To assess the significance of the reported correlations between measurements in a manner reasonably robust to outliers, we used the bootstrap method to approximate the sampling distribution of each correlation coefficient  $r$ . The data set was resampled with replacement  $10^4$  times, and for each resampling the pairwise correlations were recomputed. Positive (or negative) correlations were deemed significant if  $r = 0$  fell below the 5th (or above the 95th) percentile of the estimated distribution of  $r$ . Differences in the mean values of each measure between the perturbed and unperturbed populations were assessed for significance with the same procedure.

35. Gander, W., Golub, G. H. & Strebel, R. Least-squares fitting of circles and ellipses. *BIT* **34**, 558–578 (1994).

36. Wilson, C. A. & Theriot, J. A. A correlation-based approach to calculate rotation and translation of moving cells. *IEEE Trans. Image Process.* **15**, 1939–1951 (2006).

# Proteasome subunit Rpn13 is a novel ubiquitin receptor

Koraljka Husnjak<sup>1,2\*</sup>, Suzanne Elsasser<sup>3\*</sup>, Naixia Zhang<sup>4\*</sup>, Xiang Chen<sup>4</sup>, Leah Randles<sup>4</sup>, Yuan Shi<sup>3</sup>, Kay Hofmann<sup>5</sup>, Kylie J. Walters<sup>4</sup>, Daniel Finley<sup>3</sup> & Ivan Dikic<sup>1,2,6</sup>

**Proteasomal receptors that recognize ubiquitin chains attached to substrates are key mediators of selective protein degradation in eukaryotes. Here we report the identification of a new ubiquitin receptor, Rpn13/ARM1, a known component of the proteasome. Rpn13 binds ubiquitin through a conserved amino-terminal region termed the pleckstrin-like receptor for ubiquitin (Pru) domain, which binds K48-linked diubiquitin with an affinity of approximately 90 nM. Like proteasomal ubiquitin receptor Rpn10/S5a, Rpn13 also binds ubiquitin-like (UBL) domains of UBL-ubiquitin-associated (UBA) proteins. In yeast, a synthetic phenotype results when specific mutations of the ubiquitin binding sites of Rpn10 and Rpn13 are combined, indicating functional linkage between these ubiquitin receptors. Because Rpn13 is also the proteasomal receptor for Uch37, a deubiquitinating enzyme, our findings suggest a coupling of chain recognition and disassembly at the proteasome.**

In eukaryotes, selective protein degradation is performed primarily by the ubiquitin–proteasome pathway. The 26S proteasome is a huge macromolecular machine that contains a proteolytically active 20S core particle capped at one or both ends by a 19S regulatory particle<sup>1</sup>. The regulatory particle recognizes ubiquitinated substrates, deconjugates ubiquitin chains and unfolds substrates before their translocation into the core particle. Proteasome subunit Rpn10/S5a was shown to bind ubiquitin chains through ubiquitin-interacting motifs (UIMs)<sup>2</sup>. Receptors were subsequently identified that are not integral proteasome subunits, but deliver ubiquitinated targets to the proteasome (for reviews, see refs 3 and 4). Canonical members of this UBL/UBA family of receptors are Rad23 (hHR23a/b in humans), Dsk2 (hPLIC-1/2 in humans) and Ddi1 (refs 5–9). UBA domains of this protein family bind ubiquitin<sup>10–12</sup>, whereas their UBL domains interact reversibly with the proteasome, principally through Rpn1, but potentially also through Rpn10 (refs 13–15).

Another interesting component of the proteasome is Rpn13/ADRM1/ARM1 (refs 16–21), which docks at the regulatory particle through an N-terminal region that binds Rpn2 (refs 18, 21–23). Its carboxy-terminal region binds deubiquitinating enzyme Uch37/UCHL5 and enhances its isopeptidase activity<sup>18,20,21</sup>. Uch37 may function as an editing isopeptidase that rescues poorly ubiquitinated substrates from being degraded<sup>24</sup>.

## A ubiquitin-interactor screen identifies Rpn13

Using a yeast two-hybrid screen, with a bait of ubiquitin lacking the last two glycines to prevent its conjugation<sup>25</sup>, we identified the N-terminal segment of human Rpn13 (hRpn13) as a ubiquitin-binding partner. The interaction was confirmed using murine Rpn13 (mRpn13) as bait against monoubiquitin and Rpn2 as prey (Fig. 1a). Rpn13 from *Saccharomyces cerevisiae* (scRpn13) aligns with the ubiquitin-binding N-terminal region of hRpn13 (Fig. 1b). Comprehensive sequence analysis using profiles and hidden Markov models failed to reveal similarity to known ubiquitin- or proteasome-binding motifs (Fig. 1c and data not shown). Deletion mutants

encompassing residues 1–150 were tested for tetraubiquitin binding, thus mapping the minimal binding domain to residues 1–130 (Fig. 1d). Although smaller fragments of mRpn13 also showed detectable binding to ubiquitin, they were unstable and expressed poorly as glutathione S-transferase (GST) fusions.

The significance of the ubiquitin–Rpn13 interaction would be supported if it were conserved from yeast to mammals, particularly as budding yeast Rpn13 is truncated and the conserved N-terminal region (Fig. 1c) is only 25% identical to mammalian forms (Fig. 1b). The existence of an unidentified ubiquitin receptor in yeast had been evident to us from the viability of an *rpn10-uim rad23Δ dsk2Δ ddi1Δ* mutant (data not shown). *rpn13Δ* mutants, which are viable but show defects in protein degradation<sup>16,26</sup>, were used to test whether Rpn13 binds ubiquitin chains in the context of intact, purified proteasomes.

## Rpn13 docks ubiquitin conjugates at the proteasome

Ubiquitin chain binding by purified proteasomes can be assayed by native gel electrophoresis<sup>5,14</sup>. Proteasomes are visualized in this system by an activity stain, using a fluorogenic peptide substrate. For wild type, the predominant proteasome species contains one regulatory particle on either end of the core particle cylinder (RP<sub>2</sub>CP). Ubiquitin chains, produced by the E2 enzyme Cdc34, bind to the proteasomes and confer reduced mobility (Fig. 2a). This shift is not dependent on UBL/UBA proteins, because the proteasomes were prepared from *rad23Δ dsk2Δ ddi1Δ* mutants. A block substitution within the UIM in Rpn10 results in attenuation of the shift, reflecting Rpn10's known ubiquitin receptor function (Fig. 2a; refs 5 and 6). However, the existence of marked residual electrophoretic retardation by added chains (lane 4) indicates the presence of at least one additional ubiquitin receptor in purified proteasomes.

Addition of conjugates to proteasomes lacking Rpn13 resulted in an electrophoretic shift comparable to that of *rpn10-uim* samples (Fig. 2a). Thus the yeast orthologue of mRpn13 is active in ubiquitin chain binding and can bind ubiquitin in the context of intact

<sup>1</sup>Institute of Biochemistry II and Cluster of Excellence Macromolecular Complexes, Goethe University, Theodor-Stern-Kai 7, D-60590 Frankfurt (Main), Germany. <sup>2</sup>Tumor Biology Program, Mediterranean Institute for Life Sciences, Mestrovicovo setaliste, 21000 Split, Croatia. <sup>3</sup>Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>5</sup>Miltenyi Biotec GmbH, Stoeckheimer Weg 1, D-50829 Koeln, Germany. <sup>6</sup>Department of Immunology, Medical School University of Split, Soltanska 2, 21000 Split, Croatia.

\*These authors contributed equally to this work.



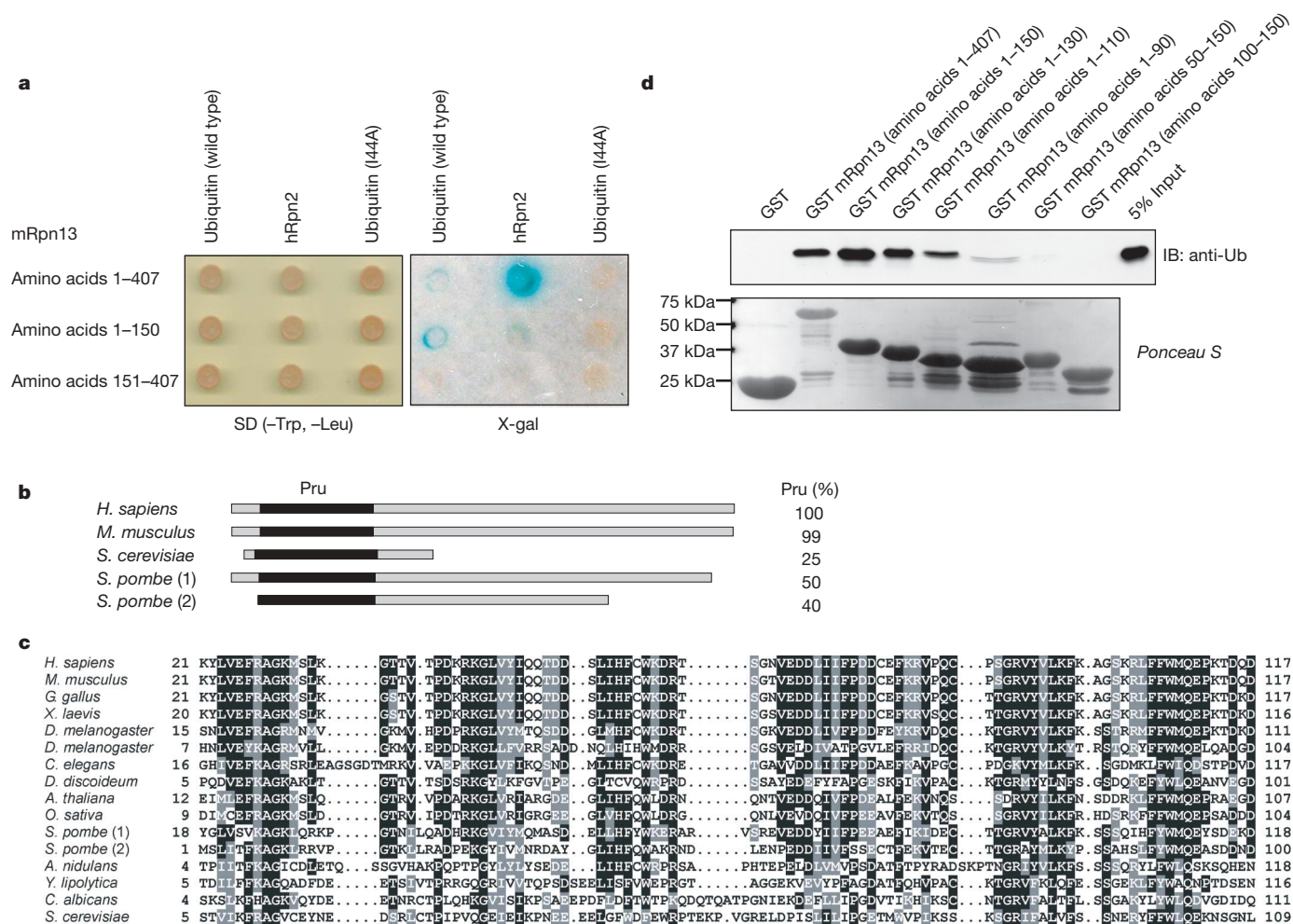
proteasomes. Indeed, its capacity for chain binding in this assay compares well with that of Rpn10. Remarkably, we observed an ostensibly complete abrogation of chain-dependent electrophoretic retardation when *rpn10-uim rpn13Δ* proteasomes were used (Fig. 2a), suggesting that Rpn10 and Rpn13 are the two major ubiquitin receptors in the yeast proteasome. However, by varying the conditions of this assay, we could, as shown below, detect residual ubiquitin chain binding with *rpn10-uim rpn13Δ* proteasomes, consistent with the existence of a still-unidentified proteasomal ubiquitin receptor. The greater abundance of the RP<sub>1</sub>CP and core particle bands in *rpn13Δ* samples suggests that Rpn13 contributes to the stability of the regulatory particle–core particle interaction *in vitro*<sup>26</sup>.

To determine whether *rpn13Δ* proteasomes were properly assembled, they were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (Fig. 2b). Apart from the absence of Rpn13 itself, the mutant proteasomes appeared to be wild type in composition. When recombinant Rpn13 was reconstituted onto mutant proteasomes after purification, their chain-binding defect was corrected (Fig. 2c). Thus, the chain-binding assay appears to report on a specific Rpn13–ubiquitin chain interaction, and not a gross structural defect of *rpn13Δ* proteasomes. GST pull-down assays also

indicated direct interaction of scRpn13 with ubiquitin (data not shown). In summary, these data indicate that Rpn13 is a novel proteasomal ubiquitin receptor.

### Loops of yeast Rpn13 bind ubiquitin

To create mutants of scRpn13 deficient in ubiquitin binding and to study the functional significance of the interaction *in vivo* subsequently, we initially used NMR to solve the structure of full-length scRpn13. These studies revealed that Thr6–Leu101 forms two contiguous, antiparallel  $\beta$ -sheets with a configuration similar to the pleckstrin-homology structural domain (Fig. 3a). In particular, a  $\beta$ -sheet comprising four antiparallel  $\beta$ -strands formed by I8–R11, E32–P37, W46–W50, and I64–L66 packs against a three-stranded sheet formed by M74–V76, I86–V90, and R96–W100. Juxtaposed to the three-stranded sheet are two  $\beta$ -strands formed by C15–N18 and L23–P26. The configuration of the  $\beta$ -strands centres around interactions between conserved aromatics within the protein core, including F10, F48, W50, W75, F87, F91, F98, F99 and W100. These findings are consistent with the crystal structure<sup>27</sup> of mRpn13. We thus named this domain pleckstrin-like receptor for ubiquitin (Pru). In a canonical pleckstrin-homology domain, K117–N130 of scRpn13



**Figure 1 | Murine Rpn13 binds ubiquitin chains.** **a**, mRpn13 cDNA fragments were cloned into pYTH9 vector in frame with the Gal4 DNA-binding domain. The resulting bait vectors were transformed into yeast strain Y190 with prey pACT2 vectors containing wild-type ubiquitin, I44A ubiquitin or hRpn2 (positive binding control) cDNA in frame with Gal4 DNA-activating domain. **b**, Architecture of Rpn13 from various species. The N-terminal domain is generally conserved (black box) whereas the C-terminal region (grey box) is absent in *S. cerevisiae* and has diverged beyond recognition in one of the two *Saccharomyces pombe* proteins (*S. pombe* (1)). *S. pombe* (1) = SPCC16A11.16; *S. pombe* (2) = SPBC342.04.

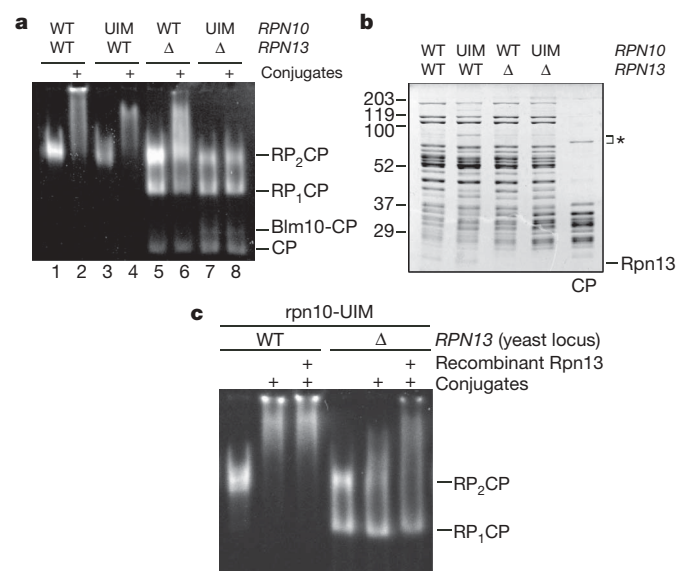
The percentage identity to the conserved hRpn13 Pru domain is provided at right. **c**, Alignment of Rpn13 N-terminal sequences. Residues that are invariant or conserved in at least 50% of sequences are shaded in black or grey, respectively. **d**, To identify the minimal region required for ubiquitin binding, mRpn13 deletion mutants were expressed as GST-fused proteins, purified and tested for their binding to linear tetraubiquitin by immunoblotting with anti-ubiquitin antibodies. Tetraubiquitin was obtained by thrombin cleavage of GST-fused tetraubiquitin (GST4×Ub) and equivalent amounts of GST-fused deletion mutants were used in GST pull-down assay.

would be  $\alpha$ -helical, consistent with secondary structure predictions for K119–G127. Residues S106–G127 are absent, however, from all acquired spectra, suggesting that this region undergoes conformational exchange and does not form a rigid helix. In the accompanying manuscript<sup>27</sup>, we find that the cognate residues in mRpn13 and hRpn13 do form helices. K117–N130 of scRpn13 shares 35.7% sequence identity with mRpn13, but the presence of a glycine at position 127 likely destabilizes the helix, as might substitution of the sequence QDE at the beginning of the helix with the more basic sequence KDK in scRpn13. Also, a salt bridge<sup>27</sup> between E119 and R122 of mammalian Rpn13 is lost, as R122 is substituted with N123 in scRpn13.

To determine how scRpn13 binds ubiquitin, we performed an NMR titration series (Supplementary Fig. 1), which implicated E41, E42, G44, F45, L66, E72, F91, S93 and R96 as being at the ubiquitin contact surface (Fig. 3b). Interestingly, these residues are in the S2–S3, S4–S5 and S6–S7 loops (Fig. 3c). The S4–S5 loop is strongly conserved in higher eukaryotes, as is F91, which is in the S6–S7 loop (Figs 1c and 3a). scRpn13 binding to monoubiquitin is in ‘fast exchange’ by NMR, which is ideal for determining binding affinity by this method, and the affinity of scRpn13 for monoubiquitin was determined to be 65  $\mu$ M (Fig. 3d).

### Rpn13 binds K48-linked diubiquitin with high affinity

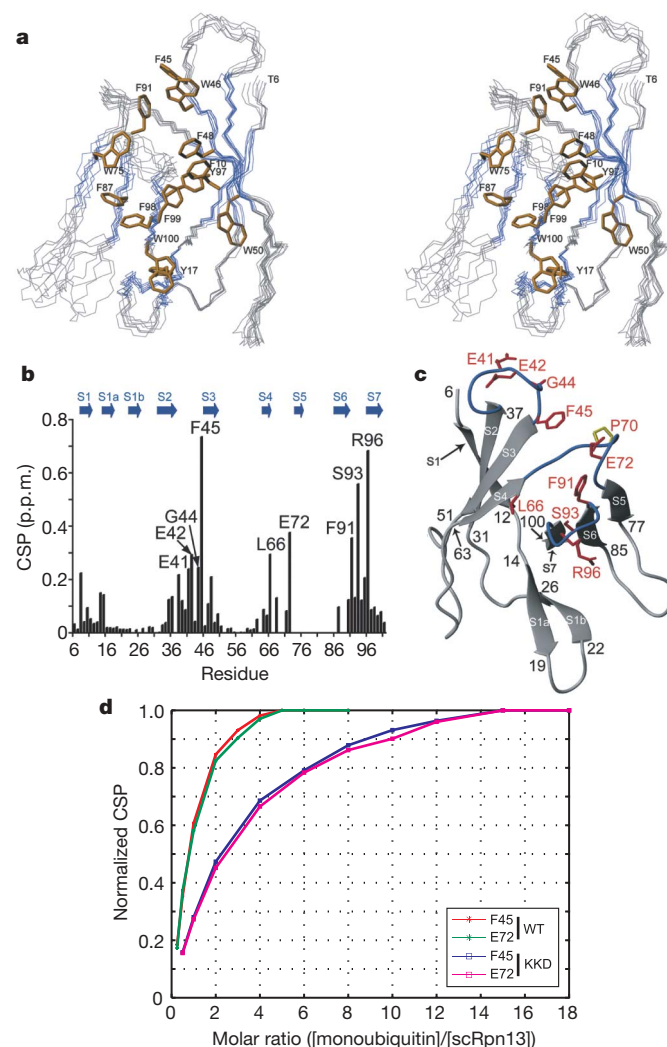
We used NMR titration experiments to determine the stoichiometry of hRpn13 for monoubiquitin, K48-linked diubiquitin and tetraubiquitin (see Supplementary Information). Monoubiquitin and diubiquitin bound to Rpn13 with 1:1 stoichiometry, whereas two Rpn13 molecules bound one tetraubiquitin (Fig. 4a, b). Therefore,



**Figure 2 | Rpn13 contributes to recognition of ubiquitin conjugates by the proteasome.** **a**, *rpn13* $\Delta$  proteasomes show defects in ubiquitin conjugate binding. Proteasomes were purified from strains (SY733, SY729, SY725 and SY722) bearing the indicated mutations. Proteasomes (4 pmol) were mixed with autoubiquitinated Cdc34 (16 pmol), resolved by native PAGE, and visualized using *N*-succinyl-Leu-Leu-Val-Tyr-(7-amino-4-methylcoumarin) (LLVY-AMC). Note that UBL/UBA proteins cannot contribute to ubiquitin chain binding in these experiments, because all proteasomes used in this figure are from a *rad23* $\Delta$  *dsk2* $\Delta$  *ddi1* $\Delta$  genetic background. *ubp6* $\Delta$  is also in the genetic background, to prevent chain disassembly during the assay. **b**, Proteasome composition is maintained in the absence of Rpn13. Proteasomes from **a** (25  $\mu$ g) were resolved by SDS-PAGE and stained with Coomassie blue. An asterisk marks contaminating protein. **c**, Reconstitution of ubiquitin conjugate binding. A subset of proteasomes from **a** (4 pmol each) was incubated with scRpn13 (20 pmol) cleaved from the GST moiety (+ lanes) or GST alone (remaining lanes) to allow reassembly, then mixed with autoubiquitinated Cdc34 (16 pmol). Complexes were resolved by native PAGE and visualized as in **a**.

although three potential diubiquitin elements exist within tetraubiquitin, no more than two Rpn13 molecules can be accommodated simultaneously. The exclusion of a third Rpn13 molecule is consistent with model structures of Rpn13–tetraubiquitin, in which steric clashes arise when three hRpn13 molecules bind neighbouring K48-linked ubiquitin subunits (Supplementary Fig. 2). That Rpn13 binds diubiquitin elements of K48-linked chains is further validated in the accompanying manuscript<sup>27</sup>.

In contrast to scRpn13, resonances broaden and shift as hRpn13 Pru binds monoubiquitin (Supplementary Fig. 3). This behaviour is associated with stronger  $K_d$  values, but prohibits their accurate calculation by the method used to determine the scRpn13–ubiquitin affinity. Fluorescence spectrophotometry was therefore used to



**Figure 3 | Rpn13 uses loops to bind ubiquitin.** **a**, Stereo view of scRpn13, spanning residues T6–L101, in which  $\beta$ -strands are indicated in blue and hydrophobic side chains in yellow. **b**, NMR titration experiments reveal scRpn13 residues that contact ubiquitin. The data were prepared as described in Methods and plotted. **c**, ScRpn13 residues that bind ubiquitin are within the S2–S3, S4–S5 and S6–S7 loops. Residues significantly affected by ubiquitin addition are displayed and labelled in red with their secondary structures in blue. **d**, scRpn13–KKD affinity for monoubiquitin is significantly reduced compared with wild type. Normalized chemical-shift perturbation values are plotted against molar ratios of monoubiquitin to wild-type scRpn13 (WT, red, green) or monoubiquitin to scRpn13–KKD (KKD, blue, purple). Each data line represents a specific amino acid, namely F45 (red and blue) and E72 (green and purple). Using Matlab version 7.2, the data were fitted to determine a binding constant of 65  $\mu$ M for wild-type scRpn13 and an eightfold reduction in the affinity of scRpn13–KKD for monoubiquitin.



determine hRpn13's affinity for monoubiquitin and diubiquitin, as hRpn13 (residues 1–150) contains two tryptophan residues and ubiquitin none. hRpn13 showed a surprisingly high affinity, with a  $K_d$  value for monoubiquitin of about 300 nM (Fig. 4c, d) and for diubiquitin of about 90 nM (Fig. 4c, d). The value for diubiquitin binding is about 15-fold lower than that of hHR23a for tetraubiquitin<sup>28</sup>. The higher affinity of hRpn13 for ubiquitin, compared with scRpn13, reflects amino-acid substitutions at the contact surface. For example, in the accompanying paper<sup>27</sup>, residues F76 and D78 of hRpn13 were implicated in hRpn13 binding to monoubiquitin. In scRpn13, however, these residues are substituted with isoleucine and glycine, respectively.

### Rpn13 recognizes a subset of ubiquitin-like proteins

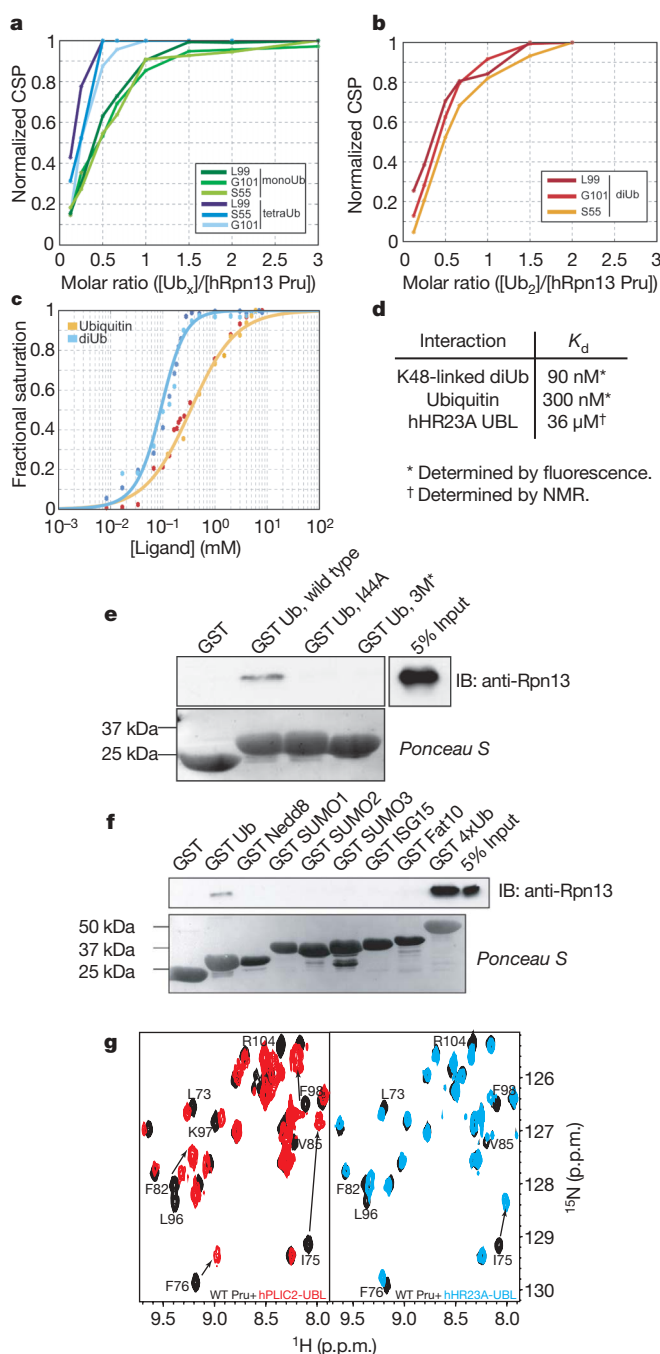
We next analysed whether Rpn13 exhibits specificity for ubiquitin or broadly recognizes ubiquitin family members. Using GST pull-down

assays, we confirmed that ubiquitin binds to full-length mRpn13 and that this interaction requires ubiquitin's hydrophobic patch, consisting of L8, I44 and V70 (Fig. 4e). mRpn13 bound more potently to linear tetraubiquitin expressed as a GST fusion (GST 4×Ub) or to purified K48-linked chains than to monoubiquitin (Fig. 4e and data not shown). Under the same experimental conditions, no binding was observed between the mRpn13 Pru domain and SUMO, Nedd8, ISG15 or FAT10 (Fig. 4f). By contrast, mRpn13 appeared to bind to the UBL domains of multiple UBL/UBA proteins (Supplementary Figs 4 and 5). We verified that hRpn13 Pru binds directly to the hHR23a and hPLIC2 UBL domains by NMR (Fig. 4g), and determined a  $K_d$  value of 36  $\mu$ M for the hRpn13 Pru–hHR23a UBL domain complex (Fig. 4d and Supplementary Fig. 5). Overlapping residues in hRpn13 were affected by the addition of these UBL domains (Fig. 4g) or ubiquitin<sup>27</sup>, suggesting that these interactions may be mutually exclusive.

### Rpn13 mutant defective in ubiquitin recognition

Experiments described above implicated residues in Rpn13's S2–S3, S4–S5 and S6–S7 loops in binding ubiquitin (Fig. 3). After introducing non-conservative substitutions for these residues individually or in combination, the resulting proteins were expressed in *Escherichia coli*, purified and characterized. We sought mutants that were defective in ubiquitin chain binding while being properly folded and proficient in proteasome binding. Separation of these functions is critical, as exemplified by previous studies of proteasomal ubiquitin receptor Rpn10. The *rpn10Δ* phenotype does not accurately reflect its function in ubiquitin recognition, because Rpn10 plays additional roles in the proteasome<sup>6,29</sup>. The proteasome is destabilized in the absence of Rpn10 (ref. 29), as is also observed in *rpn13Δ* mutants, at least under certain *in vitro* conditions (Fig. 2; ref. 26). For Rpn13, unlike Rpn10, proteasome and ubiquitin binding are conferred by the same structural domain, and thus can be effectively separated only by precisely targeted mutations. Moreover, ubiquitin contact residues in Rpn13 are distributed over three distinct loops, and thus differ from those of Rpn10 by being non-contiguous and so not subject to simple block mutagenesis.

We assayed Rpn13–proteasome binding by adding GST–Rpn13 to purified proteasomes. Because of the GST moiety, the fusion protein



**Figure 4 | Rpn13 binds to ubiquitin and UBLs of proteasomal receptors.**

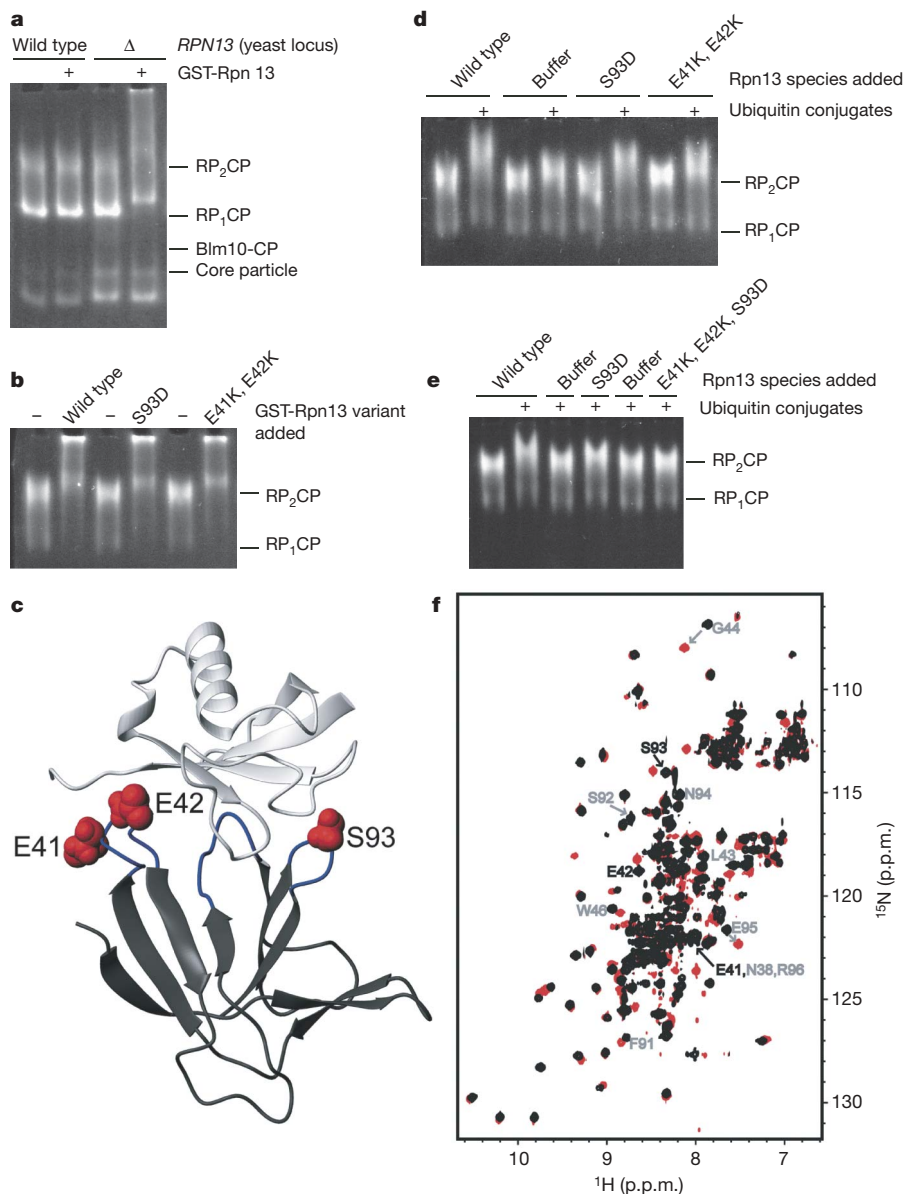
**a, b**, hRpn13 Pru binds K48-linked diubiquitin and monoubiquitin with 1:1 stoichiometry, whereas two hRpn13 Pru molecules bind one K48-linked tetraubiquitin. Normalized chemical shift perturbation values are plotted against varying molar ratios of Rpn13 to tetraubiquitin (**a**, shades of blue), diubiquitin (**b**, shades of red), or monoubiquitin (**a**, shades of green) to reveal respective Rpn13–ubiquitin binding stoichiometries of 2:1, 1:1 or 1:1. Each data line represents a specific amino acid as indicated in the figure, with values determined as described in the Supplementary Information. **c**, Binding curves for hRpn13 Pru binding to monoubiquitin or K48-linked diubiquitin as determined by intrinsic tryptophan fluorescence. Normalized fluorescence intensity values are plotted for two data sets against varying concentrations of monoubiquitin (red and orange) or diubiquitin (blue and light blue). The data were fitted by assuming 1:1 binding for hRpn13 Pru and monoubiquitin (orange) or diubiquitin (light blue). **d**, Table of hRpn13 Pru binding affinities for K48-linked diubiquitin, monoubiquitin and the UBL domain of hHR23a. Values for ubiquitin binding were determined by using the fluorescence data of (**c**). NMR titration curves were used to determine the value for the UBL domain of hHR23a. **e**, mRpn13 Pru domain was used in GST pull-down assays to assess binding to GST-tagged monoubiquitin and its mutant derivatives (I44A and triple mutant (3M\*) L8–I44–V70). **f**, mRpn13 Pru domain was used in GST pull-down assays (as in (**e**)) to assess its binding to GST-fused ubiquitin-like protein modifiers. **g**, Rpn13 binds to the hPLIC2 and hHR23A UBL domains.  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single-quantum coherence spectra of  $^{15}\text{N}$ -labelled hRpn13 Pru alone (black) and with twofold molar excess hPLIC2 (red) or hHR23A (blue) UBL domain indicates their direct interaction. Although the effect is greater for hPLIC2, these two UBL domains affect common residues in hRpn13 Pru, suggesting that they bind the same surface.



results in strong electrophoretic retardation in native gels; this effect is seen with *rpn13Δ* but not wild-type proteasomes (Fig. 5a). Thus Rpn13 assembled into the proteasome was not exchangeable with added GST–Rpn13, indicating that scRpn13 is a true proteasome subunit. Two putative ubiquitin contact-site mutants of Rpn13 were shown to be proficient in proteasome assembly (Fig. 5b). Several other mutants failed to pass this and other preliminary assays, typically because of global folding defects (data not shown). E41 and E42 are in the S2–S3 loop, and S93 in the S6–S7 loop (Fig. 5c). These sites, though greater than 22 Å apart, are both situated proximally to

bound ubiquitin in a model based on the mRpn13–monoubiquitin structure<sup>27</sup> (Fig. 5c).

To assay mutational effects on ubiquitin chain binding, we used the native gel-based assay introduced in Fig. 2a. Note that the mobility shift resulting from addition of GST–Rpn13 to the proteasome (Fig. 5a, b) is irrelevant to the chain-binding assay, because Rpn13 itself does not affect proteasome migration in gels. Only the larger GST-tagged form of Rpn13 can do so, and, in the ubiquitin chain-binding assay, untagged Rpn13 was used. Neither the S93D mutant nor the E41K, E42K mutant conferred a strong defect in the



**Figure 5 | An Rpn13 mutant specifically defective in ubiquitin chain binding.**

**a**, Reconstitution of proteasomes with recombinant GST–Rpn13. Proteasomes were purified from strains containing or lacking Rpn13 (SY775 and SY723). GST–Rpn13 (40 pmol) or buffer was mixed with proteasomes (5 pmol), which were resolved on native PAGE and visualized using suc–LLVY–AMC. The mobility shift caused by GST–Rpn13 is an indicator of its assembly into proteasomes. The presence of GST on Rpn13 is required to cause a mobility shift. All proteasomes used in this figure are from an *rpn10-uim ubp6Δ* genetic background. **b**, Mutations in Rpn13 do not attenuate assembly of Rpn13 into the proteasome. Reconstitution assays were performed as in **a**, but using a fourfold molar excess of GST–Rpn13. **c**, Structural model with mutations. Mutated residues (see **d–f**) were mapped onto a model structure of scRpn13 (dark grey) complexed with

monoubiquitin (light grey). E41, E42 and S93 are displayed in red, ubiquitin-binding loops in blue. These residues map to the S2–S3 (E41K, E42K) and S6–S7 (S93D) loops. **d**, Mutations in single loops of Rpn13 attenuate its proteasomal ubiquitin receptor function. Rpn13 variants (12 pmol) cleaved from GST were incubated with proteasomes (3 pmol) to allow reassembly. Autoubiquitinated Cdc34 (18 pmol) was then added. After 15 min at 30 °C, complexes were resolved by native PAGE and visualized using suc–LLVY–AMC. **e**, Rpn13 mutant E41K, E42K, S93D (Rpn13-KKD) abrogates the ubiquitin receptor activity of Rpn13. Experiment performed as in **d**. **f**, Superposition of <sup>1</sup>H, <sup>15</sup>N heteronuclear single-quantum coherence spectra of wild-type Rpn13 (black) and Rpn13-KKD (red). Shifted resonances are labelled in grey, and those corresponding to E41, E42 and S93 in black. Chemical shift assignments are only available for the wild-type protein.

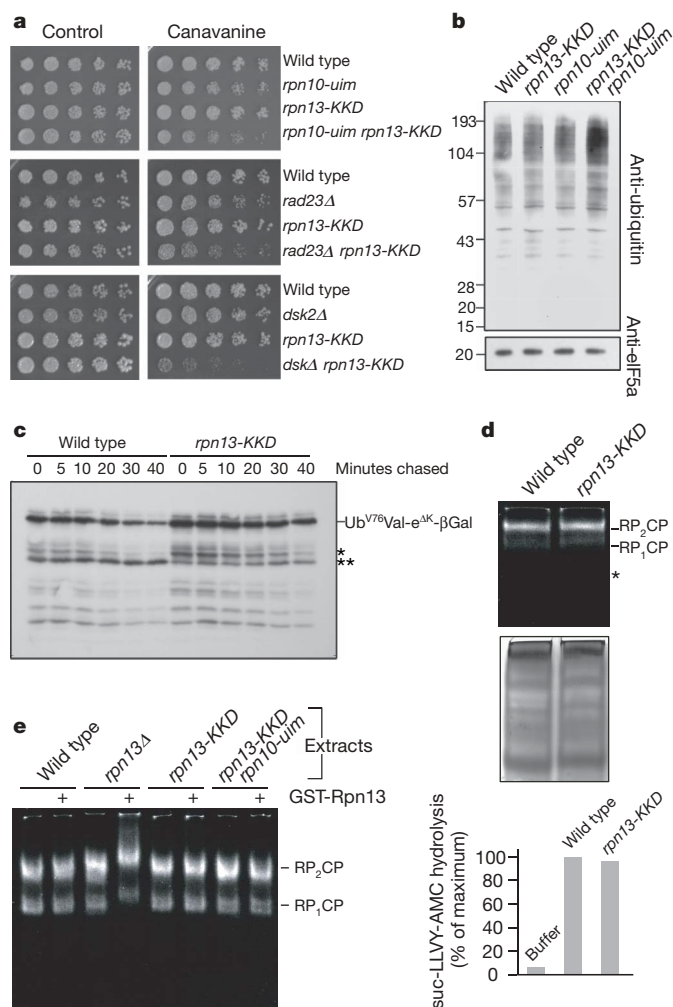
proteasome-ubiquitin interaction, although both conferred reduced mobility shifts compared with wild-type Rpn13 (Fig. 5d). To impair ubiquitin binding further, we combined the S2–S3 and S6–S7 mutations. The resulting protein, a E41K, E42K, S93D triple mutant, referred to as scRpn13-KKD, was comparable to a buffer control in its influence on the proteasome's electrophoretic mobility in the presence of ubiquitin conjugates (Fig. 5e). NMR titrations revealed that scRpn13-KKD binds monoubiquitin approximately eightfold more weakly than wild type (Fig. 3d; see also Supplementary Information). To ensure that these mutations did not affect Rpn13 structural integrity, we compared a  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single-quantum coherence spectrum recorded on  $^{15}\text{N}$ -labelled scRpn13-KKD with that of wild type (Fig. 5f). Only resonances corresponding to the mutated residues or their immediate neighbours were shifted, indicating that these surface mutations did not affect Rpn13's structure. In addition, we identified the hRpn13 surface that binds Rpn2, which is remote from the substituted residues<sup>27</sup>. The corresponding surface in scRpn13 is preserved in Rpn13-KKD as none of the residues within it were affected. Thus, Rpn13-KKD appeared to be suitable for *in vivo* analysis of the physiological function of ubiquitin recognition by Rpn13.

### Phenotype of the Rpn13-ubiquitin binding-site mutant

To test the biological significance of the Rpn13-ubiquitin interaction, we integrated the triple mutant allele into yeast in place of the wild-type chromosomal sequence. Functional defects in proteasomes can be revealed by plate assays such as sensitivity to the arginine analogue canavanine, whose incorporation into proteins causes misfolding and accelerated degradation. The enhanced substrate load

is lethal to mutants lacking proper proteasome function (see, for example, ref. 5). *rpn13-KKD* mutants proved sensitive to  $8\ \mu\text{g ml}^{-1}$  of canavanine when in the genetic background of an *rpn10-uim* mutation (Fig. 6a). Thus, the *rpn13-KKD* mutation leads to a defect in proteasome function, and interacts synthetically with another specifically targeted proteasomal ubiquitin receptor mutation. Because Rpn13 can bind UBL/UBA proteins (Fig. 4g), we also investigated its genetic relationship with Dsk2 and Rad23. *rpn13-KKD* also showed a strong synthetic interaction with a null allele of proteasomal ubiquitin receptor Dsk2 (Fig. 6a). In the case of Rad23, the genetic interaction was comparatively modest. These data support the view that the docking of ubiquitin conjugates at the proteasome by UBL/UBA proteins is not mediated obligatorily by Rpn13. In addition, binding assays performed with purified proteasomes and the UBL domains of Rad23 and Dsk2 indicate that Rpn13 is not the sole receptor for UBL/UBA proteins on the proteasome (Supplementary Fig. 7). The results of the binding assays are consistent with our previous report that proteasome subunit Rpn1 binds UBL/UBA proteins<sup>14</sup>. Further work is required to define more precisely the extent to which Rpn13-dependent docking of ubiquitin conjugates at the proteasome is mediated or regulated by UBL/UBA proteins.

To test whether amino-acid substitutions in the ubiquitin-binding loops of Rpn13 can lead to global defects in ubiquitin metabolism, whole-cell extracts from our mutants were examined by immunoblotting. High molecular mass ubiquitin conjugates, which are enriched in proteasome substrates, accumulated in the *rpn13-KKD rpn10-uim* double mutant (Fig. 6b). The defect is synthetic, as with the canavanine sensitivity of the double mutant (Fig. 6a). We also observed an *in vivo* proteolytic defect in the *rpn13-KKD* mutant (Fig. 6c), using the model proteasome substrate  $\text{Ub}^{\text{V76}}\text{-Val-e}^{\text{AK}}\text{-}\beta\text{-gal}$  (ref. 30), which was previously found to be stabilized in an *rpn13*Δ mutant<sup>16</sup>.



**Figure 6 | Phenotypic effects of the loss of ubiquitin receptor function by Rpn13.**

**a**, Canavanine sensitivity of single and double mutants in ubiquitin receptor genes. Cells in late log phase (top: SY998b, SY980f, SY1004c and SY920b; middle: SY1076, SY1073a, SY1012a and SY1080a; bottom: SY1076, SY1074a, SY1012a and SY1082a) were serially diluted and stamped on plates using a pin array. Plates were incubated at 30 °C for either 2 (left) or 3 (right) days. **b**, Endogenous ubiquitin conjugate levels in proteasomal ubiquitin receptor mutants. Cells (SY998a, SY980a, SY1004a and SY920a) were grown to log phase, and whole-cell extracts prepared. Proteins were resolved by 4–12% gradient SDS-PAGE, transferred to polyvinylidene fluoride, and probed with antibody against ubiquitin. The membrane was stripped and probed with antibody against eIF5a. **c**, Substrate stabilization in *rpn13-KKD* mutants. Cells (SY992b, SY1004b) expressing  $\text{Ub}^{\text{V76}}\text{-Val-e}^{\text{AK}}\text{-}\beta\text{-gal}$  from a GAL promoter were grown to mid-log phase under inducing conditions. Protein synthesis was quenched at time zero by adding cycloheximide. Aliquots were withdrawn at the time points indicated, and lysates prepared. Proteins were visualized by SDS-PAGE/immunoblot analysis, using an antibody to β-galactosidase, and quantified with imaging software (Kodak EDAS 290). The rate of degradation of  $\text{Ub}^{\text{V76}}\text{-Val-e}^{\text{AK}}\text{-}\beta\text{-gal}$  was reduced approximately twofold in the *rpn13-KKD* mutant compared with wild type. Asterisks indicate distinct β-galactosidase-derived partial breakdown products, whose relative intensities differ between wild type and mutant. **d**, *rpn13-KKD* mutants are not deficient in proteasome levels. Cells (SY992a, SY1004a) were grown and lysed as in **e**. Extract (150 μg) was resolved by native PAGE, and proteasome complexes visualized using suc-LLVY-AMC (top) and Coomassie blue as a loading control (middle). The asterisk indicates the expected position of the proteasome core particle, which is not visualized owing to low levels. Extracts were also subject to a quantitative proteasome assay, using suc-LLVY-AMC (bottom). **e**, Proteasomes from *rpn13-KKD* mutants are loaded with Rpn13-KKD protein. Cells (SY933, SY936, SY950 and SY952) were grown to late log phase at 30 °C in yeast extract ( $10\ \text{g l}^{-1}$ ), peptone ( $20\ \text{g l}^{-1}$ ) and dextrose ( $20\ \text{g l}^{-1}$ ) (YPD), harvested and lysed as described (see Supplementary Information). Extract (100 μg) was incubated with 1 μg of either GST-Rpn13 (+ lanes) or GST only (samples where GST-Rpn13 is absent) on ice. Proteasome complexes were resolved by native PAGE and visualized by suc-LLVY-AMC overlay assay.

Defects in proteasome assembly have been observed in *rpn13Δ* proteasomes (Fig. 2; ref. 26), and could potentially account for the phenotypes observed in Fig. 6a–c. We therefore analysed the assembly state of *rpn13-KKD* proteasomes by running native gels on freshly prepared, unfractionated cell extracts<sup>31</sup>. We observed no significant change in level, assembly or peptidase activity in *rpn13-KKD* proteasomes (Fig. 6d). Surprisingly, no assembly defect was observed for *rpn13Δ* proteasomes (Fig. 6e). Thus, assembly defects previously observed for *rpn13Δ* proteasomes are apparently a result of *in vitro* handling.

Although the recombinant Rpn13-KKD protein is properly folded and assembled onto purified proteasomes *in vitro*, it remained possible that the mutant protein would be absent from proteasomes *in vivo* as a result of its being rapidly degraded. To assess the extent of Rpn13-KKD loading of endogenous proteasomes, we used the GST–Rpn13 add-back assay of Fig. 5a, where purified proteasomes were used. In unfractionated cell extracts, GST–Rpn13 similarly shifted *rpn13Δ* proteasomes (Fig. 6e). *rpn13-KKD* proteasomes behaved as wild type in this assay, indicating that they were fully loaded with Rpn13-KKD. We conclude that the phenotypes of the *rpn13-KKD* mutant do not reflect deficient proteasome assembly or deficient loading of Rpn13 onto proteasomes, but are specifically related to its impaired ubiquitin-binding site.

## Discussion

We report here the identification of a new ubiquitin receptor for the proteasome, Rpn13, which is unrelated to Rpn10 and the three UBL/UBA proteins. Moreover, Rpn13 defines a new class of ubiquitin recognition surfaces, differing dramatically from other proteasomal ubiquitin receptors. First, whereas the UBL/UBA proteins (and perhaps Rpn10) have distinct ubiquitin- and proteasome-binding domains that are separated by flexible linkers, Rpn13 is docked into the proteasome through a surface that is in close spatial proximity to its ubiquitin-binding region. Thus Rpn13 may provide for a ubiquitin chain with precise positioning and polarity. Second, the UBL/UBA proteins, having a large population free of the proteasome and often multiple ubiquitin-binding domains, are better equipped than Rpn13 to capture ubiquitinated substrates and then deliver them to the proteasome. Third, UBL/UBA proteins are also capable of protecting the chain during transit to the proteasome, as they inhibit deubiquitination<sup>6,32,33</sup>. In striking contrast, Rpn13 promotes chain deubiquitination<sup>18,20,21</sup>. Binding to Rpn13 both facilitates the deubiquitinating activity of Uch37 (refs 20 and 21) and links Uch37 to the proteasome, suggesting that Rpn13 plays a major role in ubiquitin chain disassembly at the proteasome. Third, Rpn13 is exceptionally proficient at binding monoubiquitin and diubiquitin compared with other ubiquitin receptors associated with proteasome-mediated degradation. Although it is widely supposed that extended ubiquitin chains are required for degradation, contrary observations have been reported, in which monoubiquitin targets proteins to the proteasome<sup>24,34,35</sup>. Such substrates, although possibly rare, may be more strongly dependent on Rpn13 than those marked by chains.

Rpn13 resembles Rpn10 in its ability to bind ubiquitin-like domains of the UBL/UBA ubiquitin receptors (Fig. 4d, g and Supplementary Figs 4 and 5), implying that Rpn13 may recruit substrates to the proteasome directly or indirectly through UBL/UBA proteins. These observations support a hypothetical model whereby conjugates bind UBL/UBA family members, which dock them to the proteasome and pass them to the intrinsic receptors Rpn10 and Rpn13. In addition, compound complexes may be formed, in which a single conjugate is simultaneously bound by an intrinsic ubiquitin receptor and a UBL/UBA protein. Compound complexes may be favoured when longer chains are delivered, leading to more stable chain–proteasome interactions, and thus more rapid substrate degradation.

Ubiquitin and the proteasome are both essential, but remarkably the inactivation of all five known proteasomal ubiquitin receptors in the same yeast strain does not appear to be lethal (data not shown). In our assays, the residual chain-binding capacity of *rpn10-uim rpn13-KKD rad23Δ dsk2Δ ddi1Δ* proteasomes is modest (Fig. 5e), suggesting that the highest-affinity intrinsic receptors are now known. The unidentified receptor may be of lower affinity but comparable functionality, or it may be a receptor of typical affinity that is not intrinsic to the proteasome, such as Rad23/hHR23.

The phenotypic properties of multiply receptor-deficient strains suggest functional redundancy (Fig. 6; refs 5 and 32). This may reflect robustness in their principal function of delivering substrate to the proteasome. Given the number of known receptors, it is likely that docking of the chain at any of multiple locations in the proteasome will support breakdown of the target protein. However, there is apparently a deeper and more interesting functional relationship among ubiquitin receptors as well, in which they play distinct roles. For example, they have various phenotypes in isolation, albeit not lethal ones. Furthermore, proteasome function appears substantially compromised in multiple receptor-deficient strains (refs 5 and 32, and S.E. and D.F., unpublished observations). When different receptor mutants are compared, the relative strengths of the degradation defects vary from substrate to substrate<sup>6</sup>. Thus, the receptors show *in vivo* specificity, although it remains unclear how specific they are and what the mechanistic basis of this specificity is. Finally, functionally redundant behaviour as inferred from mutant phenotypes may not reflect the functioning of the wild-type system in a straightforward manner, because compensation can mask differentiated activities. With more detailed characterization, the individuality of proteasomal ubiquitin receptors and its mechanistic basis should become clearer.

## METHODS SUMMARY

**Yeast genetics and two-hybrid screen.** Standard methods were used for yeast genetics, growth assays and protein turnover assays (see Supplementary Information). A complete list of yeast strains is given in the Supplementary Information. Sequences corresponding to mouse ubiquitin lacking two terminal glycines (UbΔGG) were subcloned into pYTH9 vector<sup>36</sup>, creating fusion proteins with the Gal4 DNA-binding domain (bait). Yeast strain Y190 was transformed with bait vector, and human fetal brain complementary DNA (cDNA) library (Clontech) was screened as previously described<sup>36</sup>.

**Antibodies and plasmids.** Antibodies used were: anti-myc (9E10) and anti-Ub (P4D1) from Santa Cruz Biotechnology; anti-ADRM1 (anti-Rpn13) from Biomol; and anti-β-galactosidase (Promega). All constructs used in this study are described in Supplementary Information.

**Protein purification and biochemical assays.** Recombinant proteins were expressed in and purified from Rosetta cells (Novagen). Proteasome was affinity-purified essentially as described<sup>37</sup>. Immunoprecipitation, immunoblotting and GST pull-down assays were performed as previously described<sup>25</sup>. Native gel analysis was performed as in ref. 5. More detailed descriptions are available in Supplementary Information.

**NMR spectroscopy.** We determined the structure of full-length scRpn13 as described in Supplementary Information, with the data summarized in Supplementary Table 1. The resulting structures are available through Protein Data Bank accession number 2Z4D. Binding surfaces and affinities were determined as described in Supplementary Information.

Received 31 October 2007; accepted 19 March 2008.

1. Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68**, 1015–1068 (1999).
2. Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059–7061 (1994).
3. Elsasser, S. & Finley, D. Delivery of ubiquitinated substrates to protein-unfolding machines. *Nature Cell Biol.* **7**, 742–749 (2005).
4. Madura, K. Rad23 and Rpn10: perennial wallflowers join the melee. *Trends Biochem. Sci.* **29**, 637–640 (2004).
5. Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J. & Finley, D. Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. *J. Biol. Chem.* **279**, 26817–26822 (2004).



6. Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell* **118**, 99–110 (2004).
7. Kleijnen, M. F. *et al.* The hPLIC proteins may provide a link between the ubiquitination machinery and the proteasome. *Mol. Cell* **6**, 409–419 (2000).
8. Chen, L. & Madura, K. Rad23 promotes the targeting of proteolytic substrates to the proteasome. *Mol. Cell. Biol.* **22**, 4902–4913 (2002).
9. Kaplun, L. *et al.* The DNA damage-inducible UbL-UbA protein Ddi1 participates in Mec1-mediated degradation of Ho endonuclease. *Mol. Cell. Biol.* **25**, 5355–5362 (2005).
10. Bertolaet, B. L. *et al.* UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nature Struct. Biol.* **8**, 417–422 (2001).
11. Wilkinson, C. R. *et al.* Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nature Cell Biol.* **3**, 939–943 (2001).
12. Wang, Q., Goh, A. M., Howley, P. M. & Walters, K. J. Ubiquitin recognition by the DNA repair protein hHR23a. *Biochemistry* **42**, 13529–13535 (2003).
13. Hiyama, H. *et al.* Interaction of hHR23 with S5a. The ubiquitin-like domain of hHR23 mediates interaction with S5a subunit of 26 S proteasome. *J. Biol. Chem.* **274**, 28019–28025 (1999).
14. Elsasser, S. *et al.* Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nature Cell Biol.* **4**, 725–730 (2002).
15. Walters, K. J., Kleijnen, M. F., Goh, A. M., Wagner, G. & Howley, P. M. Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. *Biochemistry* **41**, 1767–1777 (2002).
16. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
17. Sone, T., Saeki, Y., Toh-e, A. & Yokosawa, H. Sem1p is a novel subunit of the 26 S proteasome from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 28807–28816 (2004).
18. Hamazaki, J. *et al.* A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J.* **25**, 4524–4536 (2006).
19. Jorgensen, J. P. *et al.* Adrm1, a putative cell adhesion regulating protein, is a novel proteasome-associated factor. *J. Mol. Biol.* **360**, 1043–1052 (2006).
20. Qiu, X. B. *et al.* hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *EMBO J.* **25**, 5742–5753 (2006).
21. Yao, T. *et al.* Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nature Cell Biol.* **8**, 994–1002 (2006).
22. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
23. Gandhi, T. K. *et al.* Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nature Genet.* **38**, 285–293 (2006).
24. Lam, Y. A., Xu, W., DeMartino, G. N. & Cohen, R. E. Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature* **385**, 737–740 (1997).
25. Bienko, M. *et al.* Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824 (2005).
26. Seong, K. M., Baek, J. H., Yu, M. H. & Kim, J. Rpn13p and Rpn14p are involved in the recognition of ubiquitinated Gcn4p by the 26S proteasome. *FEBS Lett.* **581**, 2567–2573 (2007).
27. Schreiner, P. *et al.* Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* doi:10.1038/nature06924 (this issue).
28. Raasi, S., Orlov, I., Fleming, K. G. & Pickart, C. M. Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. *J. Mol. Biol.* **341**, 1367–1379 (2004).
29. Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
30. Johnson, E. S., Bartel, B., Seufert, W. & Varshavsky, A. Ubiquitin as a degradation signal. *EMBO J.* **11**, 497–505 (1992).
31. Schmidt, M., Hanna, J., Elsasser, S. & Finley, D. Proteasome-associated proteins: regulation of a proteolytic machine. *Biol. Chem.* **386**, 725–737 (2005).
32. Saeki, Y., Saitoh, A., Toh-e, A. & Yokosawa, H. Ubiquitin-like proteins and Rpn10 play cooperative roles in ubiquitin-dependent proteolysis. *Biochem. Biophys. Res. Commun.* **293**, 986–992 (2002).
33. Raasi, S. & Pickart, C. M. Rad23 ubiquitin-associated domains (UBA) inhibit 26 S proteasome-catalyzed proteolysis by sequestering lysine 48-linked polyubiquitin chains. *J. Biol. Chem.* **278**, 8951–8959 (2003).
34. Guterman, A. & Glickman, M. H. Complementary roles for Rpn11 and Ubp6 in deubiquitination and proteolysis by the proteasome. *J. Biol. Chem.* **279**, 1729–1738 (2004).
35. Boutet, S. C., Disatnik, M. H., Chan, L. S., Iori, K. & Rando, T. A. Regulation of pax3 by proteasomal degradation of monoubiquitinated protein in skeletal muscle progenitors. *Cell* **130**, 349–362 (2007).
36. Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y. & Dikic, I. Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* **416**, 183–187 (2002).
37. Leggett, D. S. *et al.* Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of our laboratories, as well as D. Hoeller, G. Dittmar, J. Lipscomb and M. Schmidt, for discussions, comments and reading of the manuscript. We also thank the University of Minnesota's NMR facility, Minnesota Supercomputing Institutes's Basic Sciences Computing Laboratory and E. Arriaga for allowing us to use his spectrofluorometer. We thank G. Zapart for the initial Y2H ubiquitin screening, and M. Groll and P. Schneider for allowing us to use the mRpn13–ubiquitin coordinates to generate Fig. 5c. This work was supported by grants from Deutsche Forschungsgemeinschaft (DI 931/3-1), the Cluster of Excellence 'Macromolecular Complexes' of the Goethe University Frankfurt (EXC115) (I.D.) and the National Institutes of Health (CA097004 to K.J.W.; GM043601 to D.F.; GM008700-CBITG to L.R.).

**Author Information** The structures of full-length scRpn13 are deposited in Protein Data Bank under accession number 2Z4D. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to I.D. (Ivan.Dikic@biochem2.de), D.F. (daniel\_finley@hms.harvard.edu), or K.J.W. (walte048@umn.edu).

# Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures

Zhucheng Chen<sup>1,3</sup>, Haijuan Yang<sup>1</sup> & Nikola P. Pavletich<sup>1,2</sup>

**The RecA family of ATPases mediates homologous recombination, a reaction essential for maintaining genomic integrity and for generating genetic diversity. RecA, ATP and single-stranded DNA (ssDNA) form a helical filament that binds to double-stranded DNA (dsDNA), searches for homology, and then catalyses the exchange of the complementary strand, producing a new heteroduplex. Here we have solved the crystal structures of the *Escherichia coli* RecA–ssDNA and RecA–heteroduplex filaments. They show that ssDNA and ATP bind to RecA–RecA interfaces cooperatively, explaining the ATP dependency of DNA binding. The ATP  $\gamma$ -phosphate is sensed across the RecA–RecA interface by two lysine residues that also stimulate ATP hydrolysis, providing a mechanism for DNA release. The DNA is underwound and stretched globally, but locally it adopts a B-DNA-like conformation that restricts the homology search to Watson–Crick-type base pairing. The complementary strand interacts primarily through base pairing, making heteroduplex formation strictly dependent on complementarity. The underwound, stretched filament conformation probably evolved to destabilize the donor duplex, freeing the complementary strand for homology sampling.**

Homologous recombination has a central role in the repair of DNA double-strand breaks, inter-strand crosslinks and collapsed replication forks<sup>1–3</sup>. These functions are essential for maintaining genomic integrity, and defects in recombination-mediated repair are associated with cancer<sup>3</sup>. Homologous recombination also has a key role in generating genetic diversity from bacteria to humans<sup>4</sup>.

The central reaction in recombination is the exchange of strands between two homologous DNA molecules, catalysed by the RecA family of ATPases<sup>5–7</sup> conserved from bacteria to humans. RecA binds to ssDNA in an ATP-dependent manner, forming a helical nucleoprotein filament that has  $\sim 6.2$  RecA proteins per turn and approximately three nucleotides per RecA protein<sup>8–11</sup>. The DNA is underwound and stretched, with  $\sim 18.5$  nucleotides per turn and an average rise of  $\sim 5.1$  Å per nucleotide<sup>12–14</sup>. Filament formation is highly cooperative, and its nucleation requires the binding of five to six RecA protomers<sup>15–17</sup>. The RecA–ssDNA presynaptic filament then binds to a dsDNA, forming a synaptic filament that samples for ssDNA–dsDNA homology. When homology is encountered, the ensuing strand-exchange reaction results in a postsynaptic filament where the complementary strand of the donor duplex is paired with the original ssDNA. ATP hydrolysis, which is stimulated by DNA binding, dissociates all DNA, releasing a new heteroduplex and a displaced ssDNA from the donor duplex<sup>8</sup>.

The RecA filament exists in two distinct conformations<sup>18–20</sup>. The filament formed in the presence of a non-hydrolysable ATP analogue and DNA is narrow and extended, with reported pitch values ranging from 91 to 97 Å<sup>13,14,18–20</sup>. This extended conformation is considered to be the active state that can catalyse strand exchange. The inactive filament formed in the absence of DNA is wider and compressed, with an average pitch of  $\sim 82$  Å<sup>13,14,18,19</sup>. The same two states, with similar helical parameters, have been described for eukaryotic Rad51 (refs 13, 21).

In the RecA crystal structure<sup>22</sup>, the single RecA protomer in the asymmetric unit packs along a crystallographic  $6_1$ -screw axis to form a continuous helical arrangement that resembles electron

microscopy filament reconstructions. RecA orthologues and homologues similarly form crystallographic filaments along a  $6_1$  or related screw axis<sup>9</sup>. Most crystallographic filaments of RecA resemble electron microscopy reconstructions of the inactive state, with pitch values ranging from 72 to 83 Å<sup>9</sup>. Crystallographic filaments of archaeal RadA and yeast Rad51 resemble the active state<sup>23,24</sup>, but their pitch and crystallographically imposed repeat values differ substantially from electron microscopy measurements. None of these structures contains bound DNA, however. One likely explanation is that these crystallographic filaments have altered conformations that do not precisely recapitulate the active conformation required for DNA binding<sup>25</sup>.

Assuming that the constraints of crystallizing a polymer select for altered filament conformations, we constructed *Escherichia coli* RecA–DNA complexes that represent finite segments of the filament. Using this approach, we have determined a 2.8 Å crystal structure of the active presynaptic RecA–ssDNA filament, a 3.15 Å structure of the postsynaptic filament containing the new heteroduplex DNA but lacking the displaced strand, and a 4.3 Å structure of the inactive filament.

## Overall structure of the RecA–ssDNA filament

To construct a finite segment of the filament, we fused four–six RecA genes in tandem with intervening 14-residue linkers (hereafter called RecA<sub>4</sub>, RecA<sub>5</sub> and RecA<sub>6</sub>), and mutated the first and last RecA to prevent polymerization of the fusion proteins. The DNA binding, DNA-dependent ATPase and strand-exchange activities of the RecA<sub>5</sub> and RecA<sub>6</sub> fusion proteins are comparable to those of monomeric RecA (Supplementary Figs 1 and 2).

Structures of the RecA–ssDNA presynaptic filament were obtained with both RecA<sub>5</sub> and RecA<sub>6</sub> (Supplementary Table 1). The 2.8 Å-refined RecA<sub>5</sub> structure is bound to a 15-nucleotide oligo-deoxythymidine ssDNA ((dT)<sub>15</sub>) and to five molecules of ADP-aluminium fluoride-Mg (ADP-AlF<sub>4</sub>-Mg), a non-hydrolysable ATP analogue that mimics the transition state of ATP hydrolysis.

<sup>1</sup>Structural Biology Program, <sup>2</sup>Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. <sup>3</sup>Department of Biochemistry and Structural Biology, Cornell University Weill Medical College, New York, New York 10021, USA.

These crystals contain two  $\text{RecA}_5\text{-(dT)}_{15}\text{-(ADP-AlF}_4\text{-Mg)}_5$  complexes in the asymmetric unit (Supplementary Fig. 3). The 4.3 Å-refined  $\text{RecA}_6\text{-(dT)}_{18}\text{-(ADP-AlF}_4\text{-Mg)}_6$  structure contains one complex in the asymmetric unit. The structure of the inactive filament was obtained using the  $\text{RecA}_4$  bound to four molecules of adenosine 5'-( $\beta,\gamma$ -imido)triphosphate-Mg (AMPPNP-Mg) (Supplementary Fig. 4).

The structures of the two  $\text{RecA}_5$  filaments and one  $\text{RecA}_6$  filament are essentially identical (Fig. 1a). The  $\text{RecA}_5$  filaments superimpose with a 1.03 Å root-mean-square deviation (r.m.s.d.) in the positions of 1,554 of 1,608 C $\alpha$  atoms. A corresponding  $\text{RecA}_5\text{-RecA}_6$  superposition has a 1.25 Å r.m.s.d. Thus, the active conformation of the RecA filament has a precisely conserved structure. All three complexes have an essentially straight filament axis. The helical repeat, averaged over the three filaments, is 6.16 RecA per turn (s.d. of 0.03) and the pitch is 93.96 Å (s.d. of 1.46; Supplementary Table 2).

RecA consists of a 30-residue amino-terminal  $\alpha$ - $\beta$  motif, a 240-residue  $\alpha/\beta$  ATPase core and a 64-residue globular domain (CTD)<sup>22</sup>; it binds to ssDNA using its ATPase core. As predicted<sup>18,20</sup>,  $\text{ADP-AlF}_4\text{-Mg}$  binds at the interface of two RecA protomers (Fig. 1a).

The ssDNA is located very close to the filament axis and wraps around it (Fig. 1b). The planes of the bases are approximately orthogonal to the filament axis, with their Watson-Crick edges within 1.0 Å of it, while the DNA backbone is located distally from the axis. The ssDNA is oriented with its 5' end bound to the N-terminal RecA of the fusion protein<sup>26</sup>.

The active RecA-ssDNA filament differs from previous crystallographic filaments primarily in the orientation of the recombinase relative to the filament axis, and this is associated with dissimilar RecA-RecA relationships and positions for the DNA-binding elements (Supplementary Fig. 5).

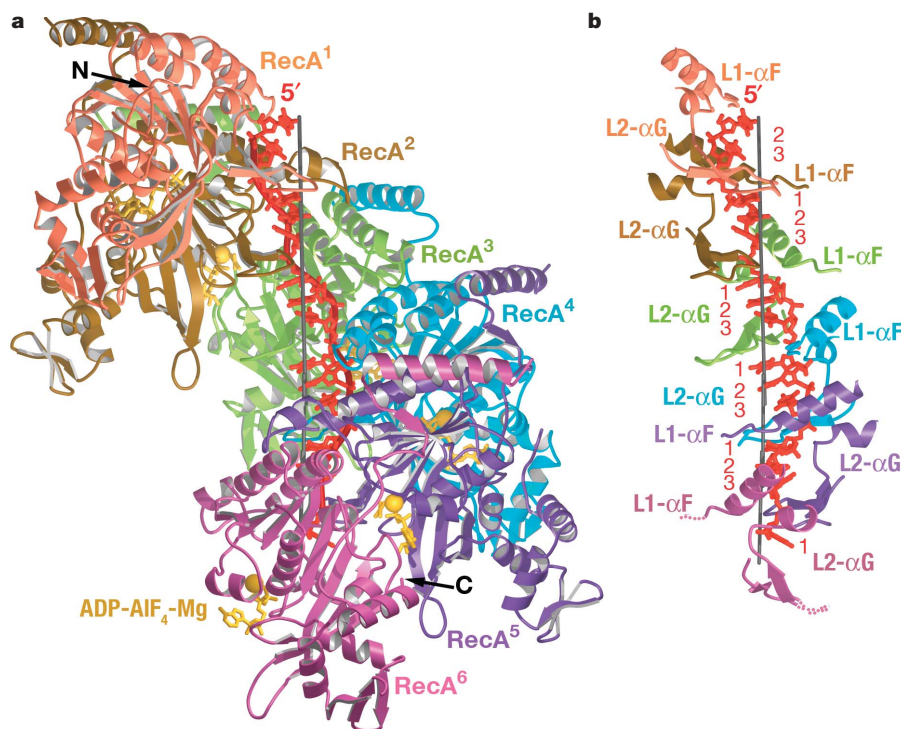
### ssDNA conformation in the presynaptic filament

The ssDNA binds with a stoichiometry of exactly three nucleotides per RecA, and the repeating unit of the DNA structure is a group of three nucleotides (hereafter called nucleotide triplet; Fig. 1b and Supplementary Fig. 6). In agreement with previous studies<sup>12,14</sup>, the overall DNA has average helical parameters of 18.5 nucleotides per turn and 5.08 Å rise per base pair. Locally, however, the bases of the nucleotide triplet are arranged in a B-DNA-like conformation, with two of the three bases stacking with a 3.5–4.2 Å spacing. In a typical nucleotide triplet, the step from the 5'-most base to the next has a 42° twist, and the next base step a 60° twist, with both steps having an axial rise of 4.2 Å and B-DNA-like slide values (Supplementary Table 3). This B-DNA-like conformation within the nucleotide triplet is compensated by the stretching (7.8 Å axial rise) and left-handed twist (−42°) of the step from the last base of one triplet to the first base of the next.

### RecA-ssDNA contacts

The ssDNA is bound by the L1 and L2 loop regions previously implicated in DNA binding<sup>10,27,28</sup>, and by the N-terminal portions of the  $\alpha$ F and  $\alpha$ G helices that follow L1 and L2, respectively (Fig. 1b). The L1 and L2 regions are disordered in the inactive filaments<sup>9</sup> but they become ordered in the presynaptic filament. L1 forms a short helix ( $\alpha$ L1) followed by a turn and an extended segment, whereas L2 forms a  $\beta$ -hairpin ( $\beta$ L2- $\beta$ 2L2).

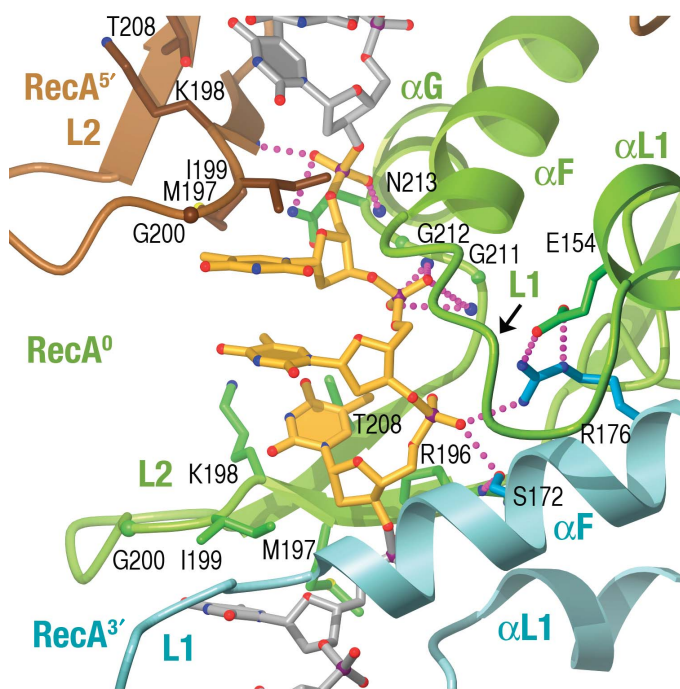
Each nucleotide triplet is bound by three consecutive RecA protomers (designated  $\text{RecA}^0$  for the RecA closest to the nucleotide triplet, and  $\text{RecA}^{5'}$  and  $\text{RecA}^{3'}$  for those of the 5'-preceding and 3'-following nucleotide triplets, respectively). The first nucleotide of the nucleotide triplet is bound by  $\text{RecA}^{5'}$  and  $\text{RecA}^0$ , the second nucleotide by  $\text{RecA}^0$ , and the third nucleotide by  $\text{RecA}^0$  and  $\text{RecA}^{3'}$ ,



**Figure 1 | Structure of the presynaptic nucleoprotein filament.** **a**, Structure of the  $\text{RecA}_6\text{-(ADP-AlF}_4\text{-Mg)}_6\text{-(dT)}_{18}$  complex. The six RecA protomers are numbered from the N-terminal RecA of the fusion protein and are coloured pink, brown, green, cyan, purple and magenta, respectively. Only 15 of the 18 nucleotides are ordered (red). The DNA backbone is traced by a red coil. The six  $\text{ADP-AlF}_4\text{-Mg}$  molecules are coloured gold. The five individual rotation/translation axes that relate adjacent RecA protomers are shown as grey

vertical lines. **b**, The L1 and L2 loop regions and the  $\alpha$ F and  $\alpha$ G helices that bind to ssDNA are coloured and numbered as in **a**, with the rest of each RecA structure omitted for clarity. The ssDNA is numbered starting with the 5'-most nucleotide in each nucleotide triplet. The 5'-most and 3'-most nucleotide triplets have only two and one ordered nucleotides, respectively. Portions of the L1 and L2 loops of the C-terminal RecA are disordered (dashed lines).





**Figure 2 | Each nucleotide triplet is bound by three consecutive RecA protomers.** The nucleotide triplet contacting RecA<sup>0</sup> is shown in yellow, whereas the previous and next triplets are in grey. RecA<sup>0</sup> is coloured green, RecA<sup>5'</sup> brown and RecA<sup>3'</sup> cyan. Hydrogen-bond interactions of the RecA<sup>0</sup> nucleotide triplet are indicated with dotted lines (magenta). See also Supplementary Fig. 7.

resulting in the strict coupling of DNA binding to a precise filament conformation (Fig. 2 and Supplementary Fig. 7).

The backbone of the nucleotide triplet is in a mostly buried environment, with all three phosphate groups bound through hydrogen bonds (Fig. 2). The 5'-most phosphate group hydrogen bonds with the backbone amide group of Met 197 from the RecA<sup>5'</sup> L2, and also with the backbone and side-chain amide groups of Asn 213 from the RecA<sup>0</sup> αG. The second phosphate group makes

two pairs of hydrogen bonds with the backbone amide groups of Gly 211 and Gly 212 from the RecA<sup>0</sup> L2. The third phosphate group hydrogen bonds with the side chains of Ser 172 and Arg 176 from the RecA<sup>3'</sup> αF.

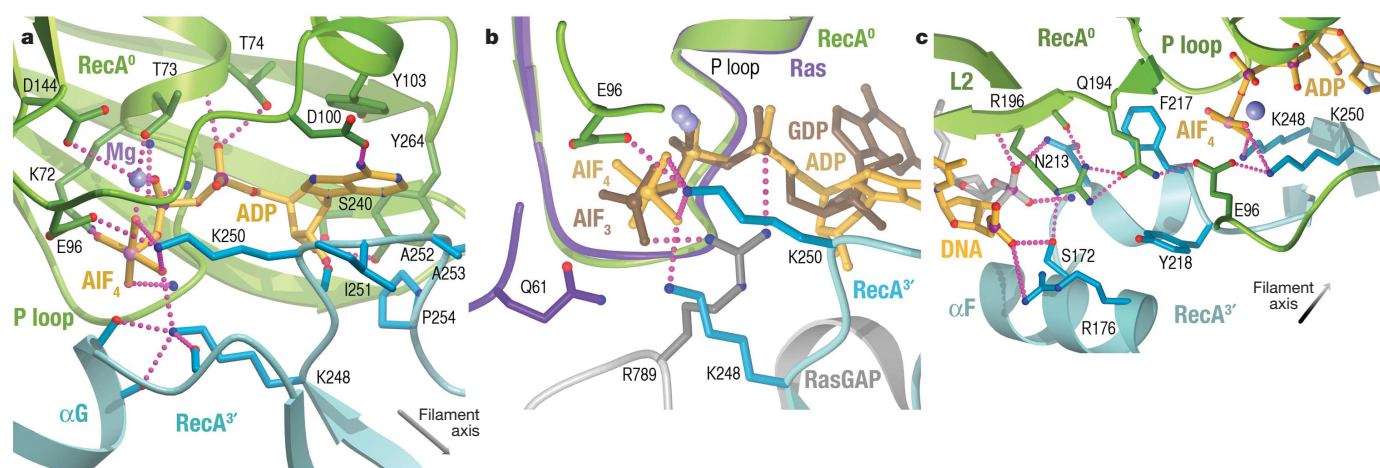
The three stacked bases are sandwiched between the L2 β-hairpins of RecA<sup>5'</sup> and RecA<sup>0</sup>, with their Watson–Crick edges solvent-exposed. The first base makes van der Waals contacts with Met 197, Ile 199 and Gly 200 from RecA<sup>5'</sup>, whereas the third base contacts aliphatic groups of Lys 198, Ile 199 and Thr 208 from RecA<sup>0</sup> (Fig. 2). These interactions help to stabilize the lack of base–base stacking in the inter-triplet gap. Ile 199 has a particularly central role, and its mutation can impair DNA repair and recombination *in vivo*<sup>29,30</sup>.

### ATP binds at a RecA–RecA interface

ADP–AlF<sub>4</sub>–Mg is sandwiched between the α/β ATPase cores of two adjacent RecA protomers in a completely buried environment (Fig. 3a). The interface of ADP–AlF<sub>4</sub>–Mg with RecA<sup>0</sup> involves the RecA<sup>0</sup> Walker motifs that coordinate the ATP phosphate groups and the Mg ion, and it is essentially the same in the inactive RecA crystal structures<sup>9</sup>.

The rest of the ADP–AlF<sub>4</sub> surface is buried at an interface with the next RecA<sup>3'</sup>. Central to this interface are the RecA<sup>3'</sup> Lys 248 and Lys 250 side chains, both of which coordinate the AlF<sub>4</sub> group (Fig. 3a). Lys 250 also hydrogen bonds to the RecA<sup>0</sup> Glu 96 side chain, which is the putative catalytic residue thought to activate a water molecule for nucleophilic attack on the γ-phosphate<sup>31</sup>. This second interface is absent in the inactive filament, where the corresponding surface of the ATP analogue is solvent exposed (Supplementary Fig. 8).

The charge-stabilized hydrogen bonds that Lys 248 and Lys 250 make to the AlF<sub>4</sub> group are expected to stabilize the active RecA–RecA interface, cooperating with DNA binding to promote the transition to the active filament state. Conversely, ATP hydrolysis and the release of inorganic phosphate would destabilize the active RecA–RecA interface by eliminating these favourable interactions and by introducing a net charge in a buried environment. Consistent with this, the K248A mutation abolishes ssDNA binding and ATP hydrolysis *in vitro*, and mutation of either Lys 248 or Lys 250 results in defective DNA repair and recombination *in vivo*<sup>31–34</sup>. In effect,



**Figure 3 | The non-hydrolysable ATP analogue ADP–AlF<sub>4</sub> binds at a RecA–RecA interface.** **a**, ADP–AlF<sub>4</sub> (yellow) binds at the interface of RecA<sup>0</sup> (green) and RecA<sup>3'</sup> (cyan). The phosphorous and aluminium atoms are coloured magenta, and the magnesium ion purple. For clarity, only a subset of the contacts between the RecA<sup>0</sup> P loop and ATP are shown. Grey arrow shows the direction of the filament axis, which is located outside the plane of view, to the left. **b**, Superposition of the RecA<sup>0</sup>–ADP–AlF<sub>4</sub>–RecA<sup>3'</sup> and Ras–GDP–AlF<sub>3</sub>–RasGAP interfaces illustrating the similarities between the RecA<sup>3'</sup> Lys 248 and Lys 250 side chains and Arg 789 of RasGAP<sup>35</sup>. RecA<sup>0</sup> and

RecA<sup>3'</sup> are coloured as in **a**, Ras is purple, RasGAP grey and GDP–AlF<sub>3</sub> brown. The catalytic side chains of RecA<sup>0</sup> (Glu 96) and Ras (Gln 61) are also shown. Orientation is similar to **a**. **c**, Intermolecular hydrogen-bond network linking the L2 loop of RecA<sup>0</sup> (green), αG of RecA<sup>3'</sup> (cyan), the RecA<sup>0</sup> nucleotide triplet (yellow), RecA<sup>3'</sup> nucleotide triplet (grey) and the AlF<sub>4</sub> group. The RecA<sup>3'</sup> Phe 217 and Tyr 218 side chains are also shown. Structural elements above the plane of the figure are omitted for clarity. The grey arrow shows the direction of the filament axis.

Lys 248 and Lys 250 act as  $\gamma$ -phosphate sensors that help establish the ATP dependency of DNA binding.

The structure reveals three mechanisms that cooperate to stimulate ATP hydrolysis. First, the interactions of Lys 248 and Lys 250 with  $\text{AlF}_4$  stabilize the transition state of ATP hydrolysis by neutralizing the charge that emerges on the  $\gamma$ -phosphate. This is analogous to the 'arginine finger' mechanism used by certain GTPase-activating proteins (GAPs). In fact, superposition of the  $\text{RecA}^0$ –ADP– $\text{AlF}_4$ – $\text{RecA}^{3'}$  and the Ras–GDP– $\text{AlF}_3$ –RasGAP<sup>35</sup> interfaces, done by aligning the P loops of  $\text{RecA}^0$  and Ras, shows that the Lys 248 and Lys 250 amino groups are positioned analogously to the guanidinium group of the RasGAP Arg 789 arginine finger (Fig. 3b). Lys 248 has been implicated in stimulating ATP hydrolysis by mutagenesis studies<sup>31</sup>.

Second, Lys 250 helps to re-orient the catalytic Glu 96 side chain, which points away from the  $\gamma$ -phosphate in the inactive RecA structures, towards the  $\gamma$ -phosphate (Fig. 3a). A similar mechanism has been proposed for the activation of the Ran GTPase by RanGAP<sup>36</sup>. Third, and again analogous to GAPs, the exclusion of solvent from the ATPase active site would further stimulate the reaction. Consistent with this, mutation of Phe 217, which together with Tyr 218 shields the catalytic Glu 96 from the solvent, decreases the ATPase activity while not substantially affecting binding to ssDNA<sup>37,38</sup> (Fig. 3c).

In essence, the 3' RecA, once correctly positioned in the active filament, activates ATP hydrolysis directly, whereas ssDNA binding, by cooperating with ATP to induce the correct positioning of the 3' RecA, acts indirectly.

### RecA–RecA interface coupled to ATP and DNA contacts

The presynaptic filament forms through two separate interfaces. The interface between the ATPase core of  $\text{RecA}^0$  and the N-terminal  $\alpha$ – $\beta$  motif of  $\text{RecA}^{3'}$  is common to both the inactive and active states. After this  $\alpha$ – $\beta$  motif, a ten-residue flexible hinge allows the  $\text{RecA}^{3'}$  ATPase core to be reoriented by a  $32^\circ$  rotation and a  $18.5 \text{ \AA}$  translation compared to the inactive filament (Supplementary Table 4 and Supplementary Fig. 8). This results in the ATPase cores of  $\text{RecA}^0$  and  $\text{RecA}^{3'}$  forming a new interface that extends continuously from the L1 and L2 loops to the ADP– $\text{AlF}_4$ . The  $\text{RecA}^0$  L1 and L2 loops now pack with  $\text{RecA}^{3'}$  and become well ordered, a transition probably aided by their DNA contacts.

A central feature of this interface is the coupling of the  $\text{RecA}^0$ – $\text{RecA}^{3'}$  interactions to RecA–DNA and RecA–ATP interactions through networks of hydrogen bonds (Fig. 3c and Supplementary Fig. 9). One of these hydrogen-bond networks starts with an interaction between the third phosphate group of the  $\text{RecA}^0$  nucleotide triplet and the  $\text{RecA}^{3'}$  Ser 172 side chain, and it sequentially links the  $\text{RecA}^0$  Arg 196 guanidinium,  $\text{RecA}^0$  Gln 194 side-chain amide,  $\text{RecA}^0$  Glu 96 carboxylate,  $\text{RecA}^{3'}$  Lys 250 amino and  $\text{RecA}^0$   $\text{AlF}_4$  groups. This network also helps to orient the catalytic Glu 96 side chain, and thus it probably contributes to the stimulation of ATP hydrolysis (Fig. 3c). Mutagenesis has implicated Gln 194, which is at the centre of this network, in the allosteric coupling of DNA binding to ATP binding<sup>29,39</sup>.

### Postsynaptic RecA filament bound to new heteroduplex

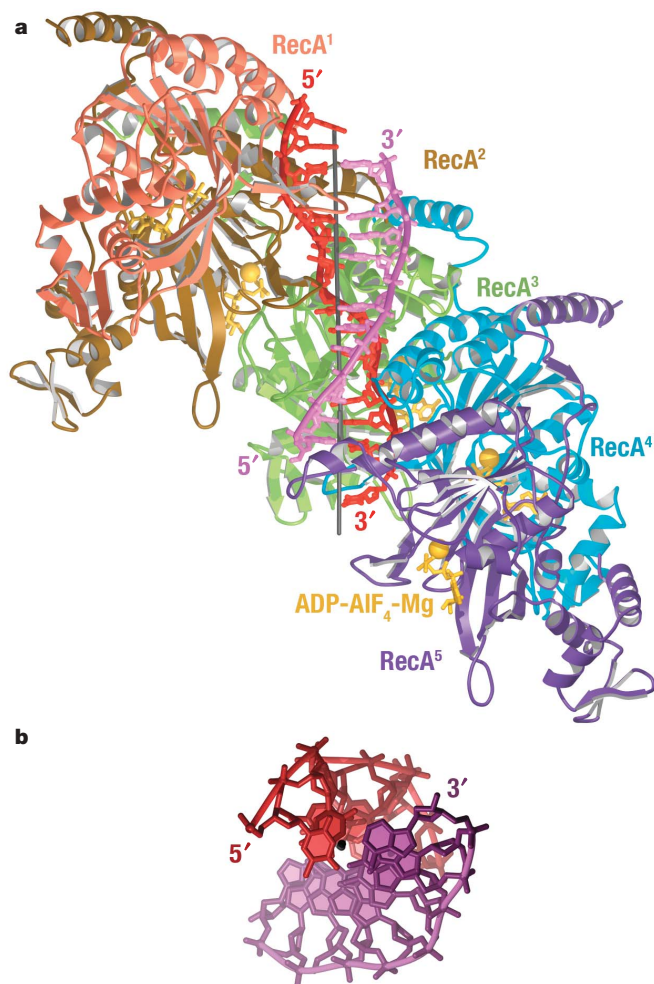
Crystals of the RecA–heteroduplex complex were obtained by incubating the  $\text{RecA}_5$ –(ADP– $\text{AlF}_4$ –Mg)<sub>5</sub> complex first with the primary d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>) strand and then with the complementary d(G<sub>2</sub>TG<sub>3</sub>) strand, followed by co-crystallization. A very similar structure was also obtained by soaking  $\text{RecA}_5$ –(ADP– $\text{AlF}_4$ –Mg)<sub>5</sub>–(dT)<sub>15</sub> crystals in a solution containing the complementary (dA)<sub>12</sub>. The structures obtained by both approaches are closely related to those of the presynaptic  $\text{RecA}_5$ –ssDNA complex, with very similar filament parameters (Fig. 4, Supplementary Table 1 and Supplementary Fig. 10).

The complementary DNA strand is juxtaposed to the primary strand in an antiparallel orientation (Fig. 4). The two strands form a duplex with a complete set of Watson–Crick hydrogen bonds (Fig. 5a). The repeating unit is now a triplet of stacked base pairs (hereafter called base-pair triplet), with adjacent base-pair triplets separated by a gap as in the presynaptic complex. The filament axis passes through the base pair plane near the major groove side (Fig. 4b).

The phosphodiester backbone of the primary strand has a very similar conformation and RecA contacts as in the presynaptic state. However, the conformations of the ribose and base groups change and become more uniform across the triplet. This allows for a more optimal Watson–Crick hydrogen-bonding geometry with the complementary strand (Fig. 5a). The complementary strand has a structure similar to that of the primary strand, except its conformation is even more regular along the triplet (Figs 4b and 5a).

Overall, the base-pair structure of the triplet closely resembles B-DNA (Fig. 5b). In a typical base-pair triplet, the first and second steps have  $3.2 \text{ \AA}$  and  $3.5 \text{ \AA}$  axial rise, and  $31^\circ$  and  $30^\circ$  twist values, respectively, with more optimal stacking compared to the presynaptic state (Supplementary Table 5). The step going from one triplet to the next still has a stretched base–base rise of  $8.4 \text{ \AA}$ , but it now has a less underwound twist of  $-4^\circ$ .

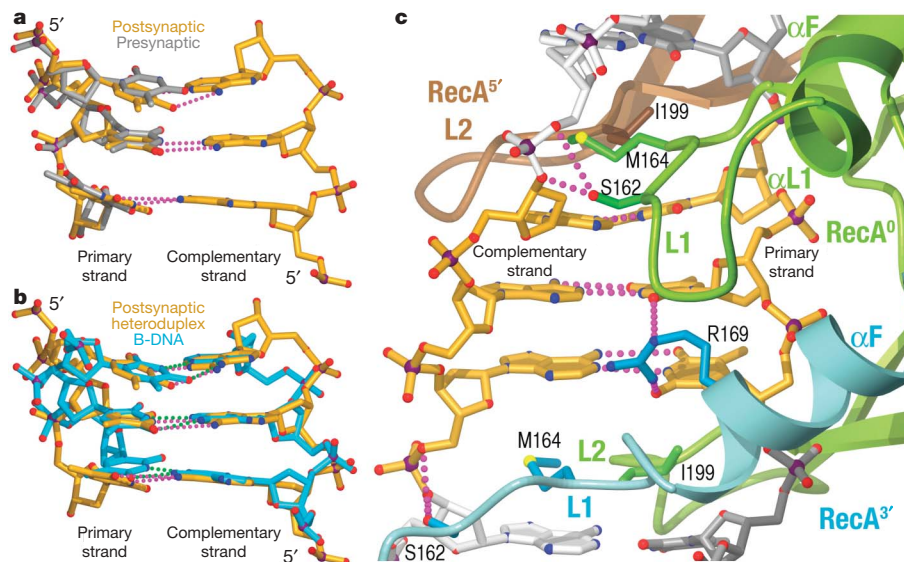
The complementary strand is held in place primarily through Watson–Crick hydrogen bonds with the primary strand, as it makes very few contacts to RecA (Fig. 5c). There is a single phosphate



**Figure 4 | Structure of the postsynaptic nucleoprotein filament.**

**a**, Structure of the  $\text{RecA}_5$ –(ADP– $\text{AlF}_4$ –Mg)<sub>5</sub>–(dT)<sub>15</sub>–(dA)<sub>12</sub> complex. The five RecA protomers are coloured as the first five protomers of Fig. 1a. The primary (dT)<sub>15</sub> strand (red) has 13 ordered nucleotides, and the complementary (dA)<sub>12</sub> strand (magenta) has 10 ordered nucleotides. **b**, View of the heteroduplex looking down the filament axis, showing the three central base-pair triplets (of  $\text{RecA}^2$ ,  $\text{RecA}^3$  and  $\text{RecA}^4$ ).





**Figure 5 | Complementary-strand binding.** **a**, Superposition of the presynaptic nucleotide triplet (grey) and postsynaptic base-pair triplet (yellow). Watson–Crick hydrogen bonds are indicated by dotted lines (magenta). The two strands of the postsynaptic base-pair triplet can be superimposed with a 1.2 Å r.m.s.d. in the positions of all phosphate and ribose atoms. **b**, Superposition of the postsynaptic base-pair triplet (yellow)

contact from Ser 162, and one set of van der Waals contacts from Met 164, both from the RecA<sup>3'</sup> L1 loop. The Met 164 side chain packs between adjacent triplets, stabilizing the inter-triplet gap in the complementary strand.

The base-pairing-induced conformational change in the primary strand is associated with a new set of base contacts. The RecA<sup>3'</sup> Arg 169 guanidinium group, which is partially disordered in the presynaptic filament, now hydrogen bonds with the O2 groups on the minor groove edges of the second and third thymidine bases. This suggests that Arg 169 may check Watson–Crick base-pairing geometry by probing the locations of the hydrogen-bond acceptor groups in the minor groove, in a manner analogous to the DNA polymerases<sup>40</sup>. Consistent with this, the R169H mutation results in ultraviolet sensitivity<sup>41</sup>.

The structure suggests that the fidelity of homologous recombination is ensured by two cooperating mechanisms. First, the limited RecA–complementary-strand contacts would make heteroduplex formation highly dependent on base pairing. Second, the B-DNA-like conformation imposed by RecA, together with the Arg 169 minor groove contacts, would exclude non-Watson–Crick hydrogen bonds, such as those in mismatched Hoogsteen or sheared base pairs. Such non-Watson–Crick base pairs with a B-DNA primary strand would result in suboptimal base–base stacking within the complementary strand.

#### Binding site for donor dsDNA and displaced strand

The RecA–ssDNA filament has a secondary site that is thought to bind to the donor dsDNA, and after strand exchange, to the displaced strand<sup>42</sup>. The Arg 243 and Lys 245 side chains implicated in this activity<sup>43</sup> are ~25 Å away from the filament axis, and because of this they have a repeat distance of ~28 Å. Modelling suggests that the binding of dsDNA to these side chains would necessitate a large separation between consecutive bases that precludes base–base stacking (Supplementary Fig. 11). This is consistent with the secondary site having weaker affinity for dsDNA compared to ssDNA<sup>42</sup>, and with micromanipulation experiments showing that stretching the dsDNA increases its affinity for RecA<sup>44</sup>.

It is thus conceivable that binding of the donor dsDNA to the secondary site, presumably through interactions limited to one of

and B-type DNA (cyan). Watson–Crick hydrogen bonds are shown as dotted lines, coloured magenta for the heteroduplex and green for B-DNA. **c**, Close-up view of the RecA–heteroduplex interface coloured as in Fig. 2, showing all the contacts to the complementary strand. The primary strand contacts that differ from the presynaptic state are also shown; the rest are omitted for clarity.

its two strands, would destabilize its double-stranded structure through loss of base stacking and pairing, releasing the complementary strand for sampling Watson–Crick hydrogen bonding with the primary strand. This model of donor–duplex destabilization could explain why the active filament has a repeat and pitch that could impose a highly underwound and stretched conformation on the DNA, even though the primary strand and heteroduplex are locally held in a B-DNA-like conformation.

#### Discussion

Our findings indicate that the RecA–ATP and RecA–DNA interactions are allosterically coupled, cooperating to induce a new RecA–RecA interface and a conformational change that activates the filament for strand exchange. The cooperativity of these interactions explains the ATP-dependency of DNA binding and the release of DNA on ATP hydrolysis.

The presynaptic ssDNA is, overall, underwound and stretched, but locally it has a conformation that is unexpectedly similar to B-DNA. This indicates that RecA catalyses strand exchange in part by holding the otherwise flexible ssDNA substrate in a conformation that resembles that of the heteroduplex. Catalysis of strand exchange would also require the destabilization of the double-stranded nature of the incoming donor duplex substrate, to free one of the donor strands for sampling base pairing with the ssDNA substrate. Our structures suggest that this may be caused by the stretching-induced disruption of base stacking and pairing when the donor duplex binds to the secondary site.

Our findings also provide insights into the fidelity of homologous recombination. The complementary strand of the heteroduplex is held in place primarily through base pairing with the primary strand. This would ensure that the strand exchange reaction is highly sensitive to base pairing, while the RecA-imposed B-DNA-like conformation would limit the base pairing to Watson–Crick-type hydrogen bonds.

#### METHODS SUMMARY

Briefly, structures of the presynaptic nucleoprotein filament were determined from RecA<sub>6</sub>–(ADP–AlF<sub>4</sub>–Mg)<sub>6</sub>–(dT)<sub>18</sub> and RecA<sub>5</sub>–(ADP–AlF<sub>4</sub>–Mg)<sub>5</sub>–(dT)<sub>15</sub> crystals. Structures of the postsynaptic filament were determined from



RecA<sub>5</sub>-(ADP-ATP<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub>-(dA)<sub>12</sub> and RecA<sub>5</sub>-(ADP-ATP<sub>4</sub>-Mg)<sub>5</sub>-d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>)-d(G<sub>2</sub>TG<sub>3</sub>) crystals, and the structure of the inactive filament from RecA<sub>4</sub>-(AMPPNP-Mg)<sub>4</sub> crystals.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 10 December 2007; accepted 4 April 2008.**

1. Luseti, S. L. & Cox, M. M. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem.* **71**, 71–100 (2002).
2. Cromie, G. A., Connelly, J. C. & Leach, D. R. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell* **8**, 1163–1174 (2001).
3. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genet.* **27**, 247–254 (2001).
4. Neale, M. J. & Keeney, S. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* **442**, 153–158 (2006).
5. Radding, C. M. Recombination activities of *E. coli* RecA protein. *Cell* **25**, 3–4 (1981).
6. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457–470 (1992).
7. Seitz, E. M., Brockman, J. P., Sandler, S. J., Clark, A. J. & Kowalczykowski, S. C. RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.* **12**, 1248–1253 (1998).
8. Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.* **3**, D570–D603 (1998).
9. Bell, C. E. Structure and mechanism of *Escherichia coli* RecA ATPase. *Mol. Microbiol.* **58**, 358–366 (2005).
10. McGrew, D. A. & Knight, K. L. Molecular design and functional organization of the RecA protein. *Crit. Rev. Biochem. Mol. Biol.* **38**, 385–432 (2003).
11. Cox, M. M. The bacterial RecA protein: Structure, function, and regulation. *Top. Curr. Gen.* **17**, 53–94 (2007).
12. Stasiak, A. & Di Capua, E. The helicity of DNA in complexes with RecA protein. *Nature* **299**, 185–186 (1982).
13. Yu, X., Jacobs, S. A., West, S. C., Ogawa, T. & Egelman, E. H. Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proc. Natl Acad. Sci. USA* **98**, 8419–8424 (2001).
14. Flory, J., Tsang, S. S. & Muniyappa, K. Isolation and visualization of active presynaptic filaments of RecA protein and single-stranded DNA. *Proc. Natl Acad. Sci. USA* **81**, 7026–7030 (1984).
15. Kelley De Zutter, J., Forget, A. L., Logan, K. M. & Knight, K. L. Phe217 regulates the transfer of allosteric information across the subunit interface of the RecA protein filament. *Structure* **9**, 47–55 (2001).
16. Galletto, R., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* **443**, 875–878 (2006).
17. Joo, C. *et al.* Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527 (2006).
18. VanLoock, M. S. *et al.* ATP-mediated conformational changes in the RecA filament. *Structure* **11**, 187–196 (2003).
19. DiCapua, E., Schnarr, M., Ruigrok, R. W., Lindner, P. & Timmins, P. A. Complexes of RecA protein in solution. A study by small angle neutron scattering. *J. Mol. Biol.* **214**, 557–570 (1990).
20. Morimatsu, K., Takahashi, M. & Norden, B. Arrangement of RecA protein in its active filament determined by polarized-light spectroscopy. *Proc. Natl Acad. Sci. USA* **99**, 11688–11693 (2002).
21. Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. Similarity of the yeast Rad51 filament to the bacterial RecA filament. *Science* **259**, 1896–1899 (1993).
22. Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* RecA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
23. Wu, Y., He, Y., Moya, I. A., Qian, X. & Luo, Y. Crystal structure of archaeal recombinase RadA: a snapshot of its extended conformation. *Mol. Cell* **15**, 423–435 (2004).
24. Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nature Struct. Mol. Biol.* **11**, 791–796 (2004).
25. Yu, X., VanLoock, M. S., Yang, S., Reese, J. T. & Egelman, E. H. What is the structure of the RecA-DNA filament? *Curr. Protein Pept. Sci.* **5**, 73–79 (2004).
26. Stasiak, A., Egelman, E. H. & Howard-Flanders, P. Structure of helical RecA-DNA complexes. III. The structural polarity of RecA filaments and functional polarity in the RecA-mediated strand exchange reaction. *J. Mol. Biol.* **202**, 659–662 (1988).
27. Wang, Y. & Adzuma, K. Differential proximity probing of two DNA binding sites in the *Escherichia coli* RecA protein using photo-cross-linking methods. *Biochemistry* **35**, 3563–3571 (1996).
28. Malkov, V. A. & Camerini-Otero, R. D. Photocross-links between single-stranded DNA and *Escherichia coli* RecA protein map to loops L1 (amino acid residues 157–164) and L2 (amino acid residues 195–209). *J. Biol. Chem.* **270**, 30230–30233 (1995).
29. Hortnagel, K. *et al.* Saturation mutagenesis of the *E. coli* RecA loop L2 homologous DNA pairing region reveals residues essential for recombination and recombinational repair. *J. Mol. Biol.* **286**, 1097–1106 (1999).
30. Larminat, F., Cazaux, C., Germanier, M. & Defais, M. New mutations in and around the L2 disordered loop of the RecA protein modulate recombination and/or coprotease activity. *J. Bacteriol.* **174**, 6264–6269 (1992).
31. Cox, J. M., Abbott, S. N., Chittineni-Pattu, S., Inman, R. B. & Cox, M. M. Complementation of one RecA protein point mutation by another. Evidence for trans catalysis of ATP hydrolysis. *J. Biol. Chem.* **281**, 12968–12975 (2006).
32. Nguyen, T. T., Muench, K. A. & Bryant, F. R. Inactivation of the RecA protein by mutation of histidine 97 or lysine 248 at the subunit interface. *J. Biol. Chem.* **268**, 3107–3113 (1993).
33. Morimatsu, K. & Horii, T. Analysis of the DNA binding site of *Escherichia coli* RecA protein. *Adv. Biophys.* **31**, 23–48 (1995).
34. Campbell, M. J. & Davis, R. W. Toxic mutations in the *recA* gene of *E. coli* prevent proper chromosome segregation. *J. Mol. Biol.* **286**, 417–435 (1999).
35. Scheffzek, K. *et al.* The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1997).
36. Daumke, O., Weyand, M., Chakrabarti, P. P., Vetter, I. R. & Wittinghofer, A. The GTPase-activating protein Rap1GAP uses a catalytic asparagine. *Nature* **429**, 197–201 (2004).
37. Skiba, M. C. & Knight, K. L. Functionally important residues at a subunit interface site in the RecA protein from *Escherichia coli*. *J. Biol. Chem.* **269**, 3823–3828 (1994).
38. Skiba, M. C., Logan, K. M. & Knight, K. L. Intersubunit proximity of residues in the RecA protein as shown by engineered disulfide cross-links. *Biochemistry* **38**, 11933–11941 (1999).
39. Kelley, J. A. & Knight, K. L. Allosteric regulation of RecA protein function is mediated by Gln194. *J. Biol. Chem.* **272**, 25778–25782 (1997).
40. Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
41. Ishimori, K. *et al.* Characterization of a mutant RecA protein that facilitates homologous genetic recombination but not recombinational DNA repair: RecA423. *J. Mol. Biol.* **264**, 696–712 (1996).
42. Mazin, A. V. & Kowalczykowski, S. C. The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *EMBO J.* **17**, 1161–1168 (1998).
43. Kurumizaka, H., Ikawa, S., Sarai, A. & Shibata, T. The mutant RecA proteins, RecAR243Q and RecAK245N, exhibit defective DNA binding in homologous pairing. *Arch. Biochem. Biophys.* **365**, 83–91 (1999).
44. Leger, J. F., Robert, J., Bourdieu, L., Chatenay, D. & Marko, J. F. RecA binding to a single double-stranded DNA molecule: a possible role of DNA conformational fluctuations. *Proc. Natl Acad. Sci. USA* **95**, 12295–12299 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. King for mass spectroscopic analysis; H. Erdjument-Bromage for N-terminal sequencing; the staff of the Advanced Photon Source ID24 beamlines for help with data collection; M. Minto for administrative assistance; and the members of the Pavletich laboratory for help and discussions. This work was supported by the NIH and the Howard Hughes Medical Institute.

**Author Information** The coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 3CMW (RecA<sub>5</sub>-(dT)<sub>15</sub>), 3CMU (RecA<sub>6</sub>-(dT)<sub>18</sub>), 3CMX (RecA<sub>5</sub>-(dT)<sub>15</sub>-(dA)<sub>12</sub>), 3CMT (RecA<sub>5</sub>-d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>)-d(G<sub>2</sub>TG<sub>3</sub>)) and 3CMV (RecA<sub>4</sub>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to N.P.P. ([pavletich@mskcc.org](mailto:pavletich@mskcc.org)).

## METHODS

**Protein engineering, expression and purification.** Multiple copies of the *Escherichia coli* *recA* gene, corresponding to amino acids 1–335, were fused into a single open reading frame with intervening 14-amino-acid linkers, each coding for the sequence Thr-Gly-Ser-Thr-Gly-Ser-Gly-Thr-Thr-Gly-Ser-Thr-Gly-Ser. The C-terminal amino acids 336–353, which are highly acidic, were omitted, as they are unstructured in other RecA crystal structures<sup>9</sup>; their truncation was previously shown to increase the DNA-binding and strand-exchange activities of RecA<sup>45</sup>. To prevent the fusion protein from polymerizing, the N-terminal RecA has a deletion of residues 1–30, which corresponds to the N-terminal  $\alpha$ - $\beta$  oligomerization motif<sup>46</sup>, and the C-terminal copy of RecA has the Cys117Met, Ser118Val and Gln119Arg mutations that disrupt oligomerization<sup>47</sup>. The fusion constructs were subcloned into a pET15b vector (Novagen) that had been modified to contain 12 histidine residues after the start codon (hereafter called His<sub>12</sub> tag) and a cleavage site for the TEV protease before the first *recA* gene.

All RecA fusion proteins were overexpressed in the recombination-deficient BLR(DE3) strain of *E. coli*. They were induced with 1 mM IPTG for 4 h at 37 °C. Cells were lysed in 50 mM Tris-HCl, 300 mM KCl, 5 mM EDTA, 5 mM dithiothreitol (DTT), 10% (v/v) glycerol, 1 mM PMSF and 1 mg ml<sup>-1</sup> each of leupeptin, aprotinin and pepstatin, pH 7.6, at 4 °C using a cell homogenizer (Avestin). After centrifugation, the DNA was precipitated from the supernatant by the addition of polyethyleneimine and NaCl, to final concentrations of 0.5% (v/v) and 0.7 M, respectively. The supernatant was then precipitated twice in lysis buffer by the addition of solid ammonium sulphate to 48% saturation. The final precipitate was dissolved in 50 mM sodium phosphate, 500 mM NaCl, pH 8.0, and it was loaded onto a Ni<sup>2+</sup> Hi-trap column (GE Health Care). After elution with 400 mM imidazole, the His<sub>12</sub> tag of the fusion protein was cleaved by incubating the eluate with 3% (w/v) TEV protease overnight, at 4 °C. The RecA fusion protein was further purified by ion exchange (Source Q and Mono Q) and gel filtration (Superdex200) chromatography. The purified protein was concentrated to 10 mg ml<sup>-1</sup> in 10 mM Tris-Cl, 200 mM NaCl, and 10 mM DTT, pH 8.0.

**Crystallization and data collection.** All crystals were grown by the hanging-drop vapour diffusion method at 16 °C. For crystallization of the presynaptic RecA-ssDNA-ADP-ALF<sub>4</sub> complexes, the RecA<sub>5</sub> or RecA<sub>6</sub> fusion protein was incubated with a threefold molar excess of ssDNA in the original protein buffer supplemented with 2 mM ADP, 10 mM MgCl<sub>2</sub> and 8 mM ALF<sub>4</sub>, pH 6.0. The ALF<sub>4</sub> solution was prepared by mixing NaF and Al(NO<sub>3</sub>)<sub>3</sub> at a 1:4 molar ratio. Crystals of the RecA<sub>6</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>6</sub>-(dT)<sub>18</sub> complex were grown from 50 mM HEPES-Na<sup>+</sup>, 1.5% (w/v) polyethylene glycol (PEG) 3350, 4% (w/v) polyvinylpyrrolidone K15 (PVP K15), 25% (v/v) 2-methyl-2,4-pentandiol (MPD), and 10 mM DTT, pH 7.5. Crystals of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub> complex were grown from 50 mM Tris-Cl, 9% (w/v) PVP K15, 32% (v/v) MPD, and 10 mM DTT, pH 8.0.

For the crystallization of the postsynaptic RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>)-d(G<sub>2</sub>TG<sub>3</sub>) complex, the RecA<sub>5</sub> fusion protein was incubated with a 1.5-fold molar excess of the primary d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>) oligonucleotide in protein buffer supplemented with 2 mM ADP, 10 mM MgCl<sub>2</sub> and 8 mM ALF<sub>4</sub>, pH 6.0, for 30 min, followed by the addition of twofold molar excess of the complementary d(G<sub>2</sub>TG<sub>3</sub>) oligonucleotide. The crystals grew from 50 mM Tris-Cl, 9% (w/v) PVP K15, 32% (v/v) MPD, 100 mM magnesium acetate, and 10 mM DTT, pH 8.0. Crystals of the postsynaptic RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub>-(dA)<sub>12</sub> complex were obtained by soaking the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub> crystals in a 0.2 mM solution of the complementary (dA)<sub>12</sub> oligonucleotide in 25 mM Tris-Cl, 9% (w/v) PVP K15, 32% (v/v) MPD, 2 mM ADP, 8 mM ALF<sub>4</sub>, and 10 mM MgCl<sub>2</sub> for 4 h.

Crystals of the inactive RecA<sub>4</sub>-(AMPPNP-Mg)<sub>4</sub> complex were obtained by pre-incubating the RecA<sub>4</sub> fusion protein with 2 mM AMPPNP and 10 mM MgCl<sub>2</sub>, pH 7.5, followed by crystallization from 100 mM Bis-Tris-Cl, 7% (w/v) PEG 3350, 100 mM KSCN, and 25% (v/v) ethylene glycol, pH 6.5. All crystals were harvested directly from the crystallization drop and flash-frozen in liquid nitrogen. Diffraction data were collected at -170 °C at the ID24C beamline of the Advanced Photon Source, and they were processed with the HKL2000 suite<sup>48</sup>.

Crystallization trials with RecA fusion proteins bound to a variety of dsDNA molecules, with and without primary ssDNA, did not yield diffraction-quality crystals. We also have not seen interpretable electron density for ssDNA at the displaced-strand/dsDNA binding site in a variety of crystallization experiments, including in the crystals of the presynaptic RecA<sub>6</sub> and RecA<sub>5</sub> crystals, which were grown in the presence of threefold molar excess of the respective ssDNA oligonucleotides. We presume that this is in part due to the weak DNA affinity of this site, and in part to the occlusion of the displaced-strand binding site, as defined by the positions of Arg 243 and Lys 245, by crystal packing contacts. These crystal

packing contacts are formed either by the non-crystallographic symmetry (NCS)-related complex in the RecA<sub>5</sub> crystal form, or by the crystallographic symmetry-related complex in the RecA<sub>6</sub> crystal form.

**Structure determination and refinement.** The first crystals obtained were those of the RecA<sub>6</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>6</sub>-(dT)<sub>18</sub> complex. They form in space group *P*<sub>3</sub>21 and have one complex in the asymmetric unit. The structure was solved by molecular replacement with the PHASER package of programs<sup>49</sup>, using the ATPase core (residues 36–269) of the monomeric RecA crystal structure (PDB code 2REB) as the search model. The model was built with the program O<sup>50</sup> and was improved by cycles of manual rebuilding and refinement with REFMAC5<sup>49</sup>. Refinement included TLS parameters for rigid bodies corresponding to the ATPase core and the C-terminal globular domains of the individual RecA protomers, and pseudo-six-fold NCS restraints across the six RecA protomers. The initial 4.3 Å model-phased *F*<sub>o</sub> - *F*<sub>c</sub> maps had 4–7  $\sigma$  electron density for the phosphate groups of the DNA, and averaging over the central four RecA protomers gave interpretable density for the DNA bases as well. The final refinement was done using the high-resolution structure of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub> complex as the starting model. The RecA<sub>6</sub> crystals exhibit anisotropic diffraction, and the overall anisotropic scale factors, calculated by REFMAC5, are *B*<sub>11</sub> = *B*<sub>22</sub> = 12.7 Å<sup>2</sup> and *B*<sub>33</sub> = -19.0 Å<sup>2</sup>. In the crystals, the packing of the RecA<sub>6</sub> protein does not result in a continuous crystallographic filament, and the 3<sub>2</sub> screw axis of the space group is coincidental.

The crystals of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub> complex form in space group *P*<sub>2</sub>1<sub>2</sub>2 and contain two complexes in the asymmetric unit. The structure was solved by molecular replacement using the central four-RecA portion of the RecA<sub>6</sub> crystal structure as a search model. It was refined with REFMAC5 using TLS groups and weak pseudo-NCS restraints over the ten RecA-nucleotide-triplet segments of the two complexes in the asymmetric unit. The overall anisotropic scale factors are *B*<sub>11</sub> = 1.5 Å<sup>2</sup>, *B*<sub>22</sub> = -0.1 Å<sup>2</sup> and *B*<sub>33</sub> = -1.4 Å<sup>2</sup>.

The crystals of the two RecA<sub>5</sub>-heteroduplex complexes are closely related to the crystals of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub> complex, but they are not isomorphous owing to small rotations and shifts in the arrangement of the two NCS-related copies. Their structures were determined by molecular replacement and refined with REFMAC5 as before, except for the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>)-d(G<sub>2</sub>TG<sub>3</sub>) structure where the NCS restraints for the DNA were applied only across the dimer in the asymmetric unit owing to the non-uniform DNA sequence of that complex.

The crystals of the inactive RecA<sub>4</sub>-(AMPPNP-Mg)<sub>4</sub> complex form in space group *P*<sub>2</sub>1<sub>2</sub>2<sub>1</sub> and contain eight complexes in the asymmetric unit. The structure was solved by molecular replacement with PHASER, using residues 37–328 of the monomeric RecA crystal structure (PDB code 2REB) as the search model. Inspection of the initial set of molecular replacement solutions allowed the identification of the RecA<sub>4</sub> fusion protein, which was then used as the search model to find the remaining complexes by molecular replacement. The structure was refined using tight pseudo-NCS restraints across the 32 RecA protomers in the asymmetric unit. No NCS restraints were applied to the relative arrangement of RecA protomers within each RecA<sub>4</sub> fusion protein, and the filament parameters of the eight RecA<sub>4</sub>-(AMPPNP-Mg)<sub>4</sub> complexes should not be influenced by refinement. The overall anisotropic scale factors are *B*<sub>11</sub> = -4.6 Å<sup>2</sup>, *B*<sub>22</sub> = 6.5 Å<sup>2</sup> and *B*<sub>33</sub> = -1.9 Å<sup>2</sup>.

The refined structures of the presynaptic and postsynaptic filaments contain residues 37–333 of the N-terminal RecA, residues 1–333 of the internal RecA protomers, and residues 1–154, 168–196 and 205–328 of the C-terminal RecA of each fusion protein. The internal disordered regions of the C-terminal RecA correspond to portions of the L1 and L2 loop regions, which presumably are disordered owing to the lack of the next RecA on which they would pack. The primary ssDNA structures typically have the 5'-most nucleotide of the first triplet and the 3'-most two nucleotides of the last triplet disordered, again probably due to the lack of adjacent RecA protomers with which they would also interact. The complementary ssDNA structure of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-(dT)<sub>15</sub>-(dA)<sub>12</sub> complex has three complete triplets and one nucleotide of the 5'-most triplet ordered, whereas that of the RecA<sub>5</sub>-(ADP-ALF<sub>4</sub>-Mg)<sub>5</sub>-d(T<sub>5</sub>C<sub>3</sub>AC<sub>2</sub>T<sub>4</sub>)-d(G<sub>2</sub>TG<sub>3</sub>) complex has all six nucleotides ordered.

The structure of the inactive RecA<sub>4</sub>-(AMPPNP-Mg)<sub>4</sub> complex contains residues 37–151, 168–194 and 211–328 of the N-terminal RecA, and residues 6–151, 168–194 and 211–328 of the remaining three RecA protomers. The internal deletions correspond to the L1 and L2 loop regions. Several of the 32 RecA protomers of the asymmetric unit have portions of the L1 or L2 loops ordered through interactions with NCS-related RecA<sub>4</sub> protomers. All DNA-bound structures contain one ADP-ALF<sub>4</sub>-Mg complex per RecA, and the inactive complex contains one AMPPNP-Mg per RecA. There are no water molecules in any of the models. The statistics from the crystallographic analyses are summarized in Supplementary Table 1.

The linkers that connect adjacent RecA protomers in the fusion proteins are unstructured. In the active filaments, the last ordered residue (333) of one RecA and the first (1) of the next RecA are separated by 25 Å, with a clear line of sight. This is approximately half the distance the 14-residue linker plus the 2 disordered RecA residues could span in an extended conformation, and it is thus unlikely that fusing the RecA protomers into a single protein caused any alterations to the filament conformation.

45. Benedict, R. C. & Kowalczykowski, S. C. Increase of the DNA strand assimilation activity of RecA protein by removal of the C terminus and structure-function studies of the resulting protein fragment. *J. Biol. Chem.* **263**, 15513–15520 (1988).
46. Mikawa, T., Masui, R., Ogawa, T., Ogawa, H. & Kuramitsu, S. N-terminal 33 amino acid residues of *Escherichia coli* RecA protein contribute to its self-assembly. *J. Mol. Biol.* **250**, 471–483 (1995).
47. Dutreix, M., Burnett, B., Bailone, A., Radding, C. M. & Devoret, R. A partially deficient mutant, RecA1730, that fails to form normal nucleoprotein filaments. *Mol. Gen. Genet.* **232**, 489–497 (1992).
48. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
49. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
50. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).



# A Lévy flight for light

Pierre Barthelemy<sup>1</sup>, Jacopo Bertolotti<sup>1</sup> & Diederik S. Wiersma<sup>1</sup>

A random walk is a stochastic process in which particles or waves travel along random trajectories. The first application of a random walk was in the description of particle motion in a fluid (brownian motion); now it is a central concept in statistical physics, describing transport phenomena such as heat, sound and light diffusion<sup>1</sup>. Lévy flights are a particular class of generalized random walk in which the step lengths during the walk are described by a 'heavy-tailed' probability distribution. They can describe all stochastic processes that are scale invariant<sup>2,3</sup>. Lévy flights have accordingly turned out to be applicable to a diverse range of fields, describing animal foraging patterns<sup>4</sup>, the distribution of human travel<sup>5</sup> and even some aspects of earthquake behaviour<sup>6</sup>. Transport based on Lévy flights has been extensively studied numerically<sup>7–9</sup>, but experimental work has been limited<sup>10,11</sup> and, to date, it has not seemed possible to observe and study Lévy transport in actual materials. For example, experimental work on heat, sound, and light diffusion is generally limited to normal, brownian, diffusion. Here we show that it is possible to engineer an optical material in which light waves perform a Lévy flight. The key parameters that determine the transport behaviour can be easily tuned, making this an ideal experimental system in which to study Lévy flights in a controlled way. The development of a material in which the diffusive transport of light is governed by Lévy statistics might even permit the development of new optical functionalities that go beyond normal light diffusion.

In recent years, light has become a tool widely used to study transport phenomena. Various analogies between electron, light and matter-wave transport have been discovered, including weak and strong localization<sup>12</sup>, the Hall effect<sup>13</sup>, Bloch oscillations<sup>14</sup> and universal conductance fluctuations<sup>15</sup>. Understanding light in disordered systems is of primary importance for applications in medical imaging (for example tumour diagnostics)<sup>16</sup>, random lasing<sup>17</sup> and image reconstruction<sup>18</sup>. Most of these studies have been limited to the simplified case in which the light performs a random walk that can be described as a diffusion process.

In a Lévy flight, the steps of the random walk process have a power-law distribution, meaning that extremely long jumps can occur<sup>2,19,20</sup> (Fig. 1). Consequently, the average step length diverges and the diffusion approximation breaks down for Lévy flights. Power-law distributions often appear in other physical phenomena that exhibit very large fluctuations, for instance the evolution of the stock market<sup>21,22</sup> and the spectral fluctuations in random lasers<sup>23,24</sup>.

A random walk in which the step length is governed by Lévy statistics leads to superdiffusion; that is, the average squared displacement  $\langle x^2 \rangle$  increases faster than linearly with time  $t$

$$\langle x^2 \rangle = Dt^\gamma$$

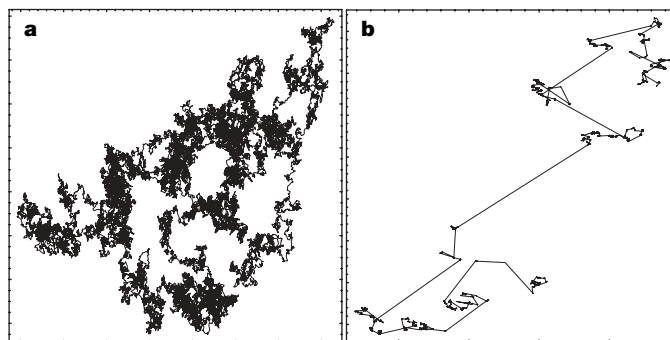
where  $\gamma$  is a parameter that characterizes the superdiffusion and  $D$  is a generalized diffusion constant. For  $\gamma > 1$  we have superdiffusion, whereas for  $\gamma = 1$  we recover normal diffusive behaviour. Normal diffusions are therefore limiting cases of Lévy flights. In Lévy flights, superdiffusion is purely a result of the long-tailed step-length

distribution. Random walks in which the step time (and thus a finite velocity) is also important are called Lévy walks<sup>19</sup>. A long-tailed distribution in the scattering dwell time can give rise to, for example, subdiffusion<sup>25</sup> ( $\gamma < 1$ ). There is no practical difference between a Lévy walk and a Lévy flight in the experiments described in this paper, because all of the experiments are static (time independent).

We report here on the creation of an optical material in which the step-length distribution can be specifically chosen. We use this to produce a structure in which light performs a Lévy flight. In a set of experiments, we show that the optical transport in such a material is superdiffusive. To produce such a structure requires an approach that initially seems counter-intuitive. The material that we have obtained is, however, relatively easy to make and provides the first well-controlled experimental test ground for Lévy transport processes. We propose the name Lévy glass for this material.

To obtain an optical Lévy flight it might seem best to develop scattering materials with self-similar (fractal) structures. This approach turns out not to work in practice, owing to the dependence of the scattering cross-section on size. In, for instance, a fractal colloidal suspension, the larger particles would be subject to resonant (Mie) scattering, whereas the smaller particles would hardly scatter at all (Rayleigh limit). The solution is to find a way to modify the density of scatterers instead of their size. This makes it possible to obtain a scattering mean free path that depends strongly on the position inside the sample.

We have found a relatively easy, but so far unstudied, method of doing this, using high-refractive-index scattering particles (of titanium dioxide in our case) in a glass matrix. The local density of scattering particles is modified by including glass microspheres of a particular, highly non-trivial size distribution. These glass microspheres do not scatter, because they are incorporated into a glass host with the same refractive index. Their sole purpose is locally to modify the density of scattering elements.



**Figure 1 | Random walk trajectories.** **a**, Normal diffusive random walk; **b**, Lévy random walk with  $\gamma = 2$  (Lévy flight). In the normal diffusive random walk, each step contributes equally to the average transport properties. In the Lévy flight, long steps are more frequent and make the dominant contribution to the transport.

<sup>1</sup>European Laboratory for Nonlinear Spectroscopy and INFN-BEC, via Nello Carrara 1, 50019 Sesto Fiorentino (Florence), Italy.

The random walk in normal diffusive materials has a gaussian step-length distribution with average step length given by the mean free path  $\ell$

$$\ell = \frac{1}{\langle N\sigma \rangle} \quad (1)$$

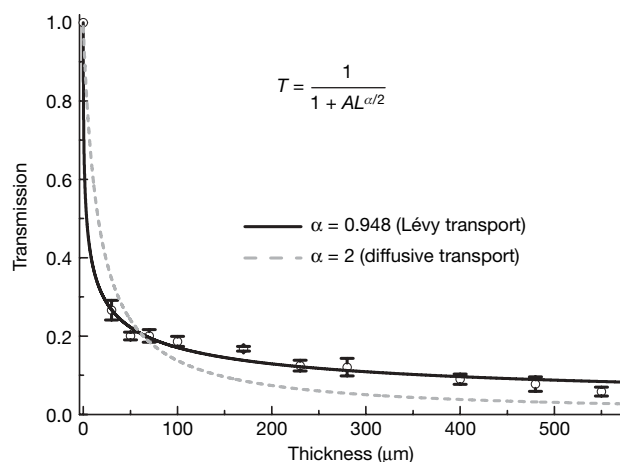
where  $\sigma$  is the scattering cross-section and  $N$  is the density of scattering elements. The angle brackets indicate an average over the sample volume. To permit Lévy flights, the material should give rise to a step-length distribution with a heavy tail, decaying as<sup>26</sup>

$$P(z) \rightarrow \frac{1}{z^{\alpha+1}}$$

where  $P(z)$  is the probability of a step of length  $z$  and  $\alpha$  is a parameter that determines the type of Lévy flight. The parameter  $\alpha$  can be shown to be related to the superdiffusion exponent  $\gamma$  by  $\gamma = 3 - \alpha$ , for  $1 \leq \alpha < 2$  (ref. 7). The moments of this distribution diverge for  $\alpha < 2$ , which means that the average in equation (1) can no longer be taken over the entire sample. However,  $N\sigma$  can still be interpreted as the local scattering strength of the material.

Our samples were made by suspending titanium dioxide nanoparticles in sodium silicate, together with a precisely chosen distribution  $P_s(d)$  of glass microspheres of different diameters  $d$ . The total concentration of titanium dioxide nanoparticles was chosen such that, on average, about one scattering event takes place in the titanium-dioxide-filled spaces between adjacent glass microspheres. The step-length distribution is then determined by the density variations induced by the distribution  $P_s(d)$  of the glass microspheres. We have calculated that a diameter distribution  $P_s(d) = 1/d^{2+\alpha}$  is required to obtain a Lévy flight with parameter  $\alpha$ , and show this experimentally below. Although with our method we can obtain a Lévy flight with any value of  $\alpha$ , we have chosen to work with  $\alpha = 1$ , because this is one of the few cases in which the Lévy distribution has a simple analytical expression (namely that of the Cauchy distribution<sup>27</sup>). For all other details on sample preparation and the derivation of the diameter distribution for Lévy flights with parameter  $\alpha$ , see Supplementary Information.

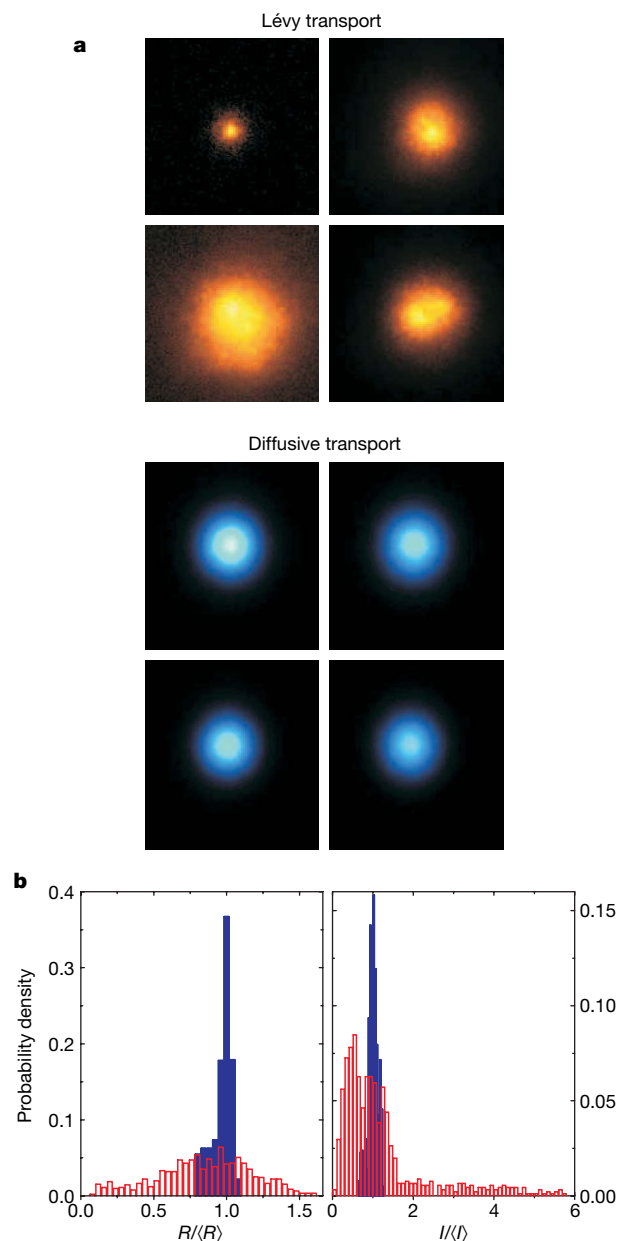
We made a series of samples of different thicknesses in the range 30–550  $\mu\text{m}$ . This allowed us to record the thickness dependence of the total transmission. To do so, a collimated He–Ne laser beam was used incident on the sample on a spot of area 1  $\text{mm}^2$ . The total transmitted light was then collected by means of an integrating sphere. Total transmission in normal diffusive systems is known to



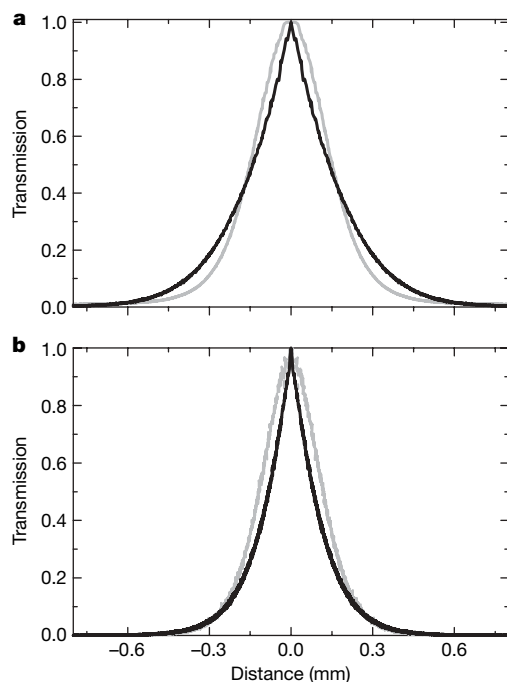
**Figure 2 | Thickness dependence of the total transmission.** For superdiffusion the transmission decays much more slowly than for normal diffusion, and should follow a power law with exponent  $\alpha/2$ . The dashed grey curve shows the normal diffusive behaviour ( $\alpha = 2$ ), whereas the black line is a fit to the data with only  $\alpha$  as free parameter. We obtain  $\alpha = 0.948 \pm 0.09$ , which is very close to the expected value,  $\alpha = 1$ , for a lorentzian Lévy flight. For very thick samples (550  $\mu\text{m}$ ), optical absorption decreases the transmission to slightly below the ideal curve. Error bars, s.d.

decay following Ohm's law, which means that the transmission depends linearly on the inverse sample thickness<sup>12</sup>. For superdiffusion this can be generalized as follows, where  $A$  is a constant and  $L$  is the thickness<sup>28</sup>:

$$T = \frac{1}{1 + AL^{\alpha/2}}$$

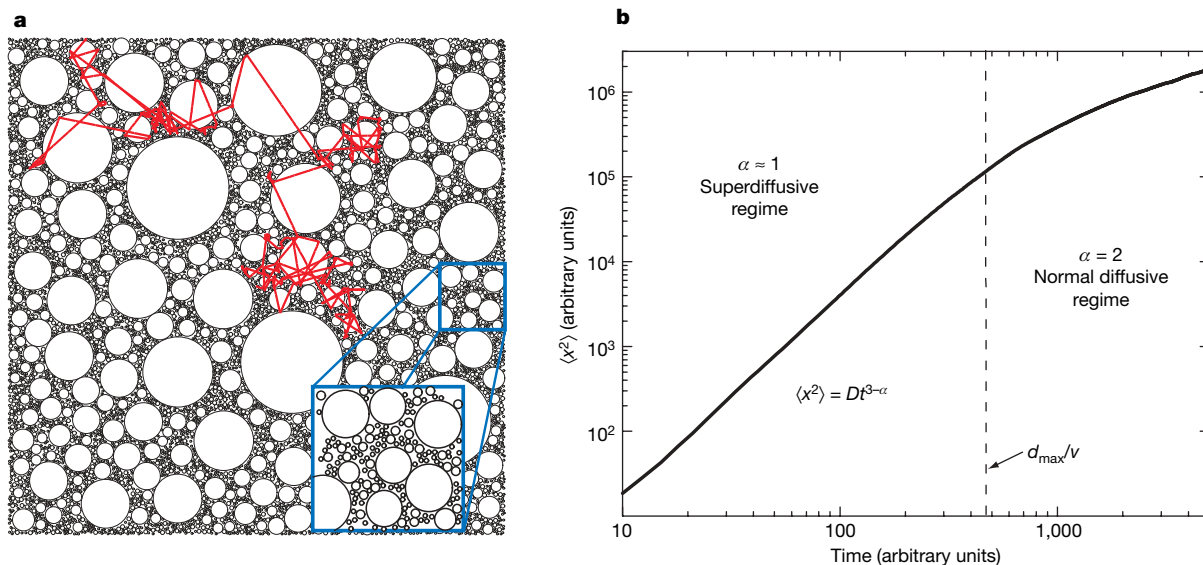


**Figure 3 | Spatial dependence of the transmission on the output surface.** **a**, Spatial distributions of the transmitted intensity for the Lévy samples (top) and for normal diffusive samples of the same thickness (bottom). The images were taken using a Peltier-cooled charged-coupled-device camera on the output surface of the sample, which was illuminated from the front with a focused (2  $\mu\text{m}$ -spot-size) He–Ne laser. The sample was placed between crossed polarizers to make sure that any residual ballistic light was blocked. The normal diffusive sample was made by using only sodium silicate and titanium dioxide powder. In the Lévy case we can see that the transmission profiles strongly fluctuate from one measurement to another, whereas in the normal diffusive case they are nearly constant. **b**, Distributions of the radius  $R$  (normalised to its average,  $\langle R \rangle$ ) and total intensity  $I$  (normalised to its average,  $\langle I \rangle$ ) of the transmission profiles for the normal diffusive (blue) and Lévy (red) samples. We can see that the very large fluctuations in the Lévy case correspond to a broad distribution function of both the intensity and radius of the transmission profile.



**Figure 4 | Average transmission on the output surface versus radial distance from the centre.** **a**, Experimental data. In the Lévy case (black) an average over 3,000 sample configurations was needed to obtain the average behaviour. The profile of the Lévy sample shows a pronounced cusp, and slowly decaying tails. The normal diffusive sample (grey) has a profile close to a gaussian lineshape: the top is rounded and long tails are absent. **b**, Result of Monte Carlo simulations of a normal diffusive random walk (grey) and a Lévy random walk (black) in a slab. The superdiffusive profile again displays a sharp cusp and decays more slowly than does the normal diffusive profile. The difference in absolute widths between experiment and simulation is due to internal reflections at the boundary of the sample, which were not taken into account in the simulations.

For the normal diffusive case in which  $\alpha = 2$ , we recover Ohm's law of conductance. The experimental data are shown in Fig. 2. We can see that they decay much more slowly than linearly, showing that transport in these samples is superdiffusive. In this case  $\alpha = 0.948 \pm 0.09$ .



**Figure 5 | Lévy walk in an inhomogeneous medium.** **a**, Random walker trajectory, obtained by Monte Carlo simulation. Owing to the strong density fluctuations, the scattering material permits Lévy flights (red). Inset, magnification showing the scale invariance of the material's structure. **b**,

This result is in excellent agreement with the expected value for a lorentzian Lévy flight, without the use of any additional fit parameters.

The power-law step-length distribution of a Lévy flight is expected to give rise to strong fluctuations in the transport properties of individual samples. In the total transmission profile we should therefore observe large differences between disorder realizations. In comparison, a normal diffusive sample would show almost no fluctuations. In Fig. 3a, we present the intensity profiles taken from the output (rear) surface of a sample that is illuminated from the front with a focused He-Ne laser. Successive images were taken by moving the sample over distances much larger than the illuminated region.

We compared the behaviour of a Lévy glass with that of a normal diffusive system of the same thickness. From the Lévy glass we observed very large differences between disorder realizations, whereas the result for the normal diffusive system is nearly constant. To quantify this behaviour we calculated the distributions of the radius and the intensity of these profiles on a set of 900 disorder realizations (Fig. 3b). In the Lévy case the distributions are extremely broad, but in the normal diffusive case they are very narrow. Moreover, in the Lévy case the distributions have slowly decaying tails, which are absent in the normal diffusive case.

The characteristics of the Lévy flight also survive if we perform an average over a large number of observations. The resulting profiles of the transmitted intensity on the output surface are plotted in Fig. 4 and compared with the results of Monte Carlo simulations. Both the experimental and the simulation results show the same features. For the normal diffusive system we observe that the profile has, as expected, a bell-shaped profile, which is very close to a gaussian curve. For the Lévy sample, however, the profile exhibits a well-defined cusp and has tails that decay much more slowly than in the normal diffusive case. The agreement between the experimental and simulated profiles is very good. The small discrepancy in the overall width of the profile can be explained by the influence of internal reflections at the boundary of the sample, which were not taken into account in the Monte Carlo simulations. We have verified that in a sample made of titanium dioxide nanoparticles and just one family of (large) glass microspheres, the profile remains gaussian (Supplementary Information). This confirms that the density variations induced by the entire size distribution of glass microspheres are required to obtain a Lévy flight.

Average squared displacement. The spreading is superdiffusive, with  $\alpha = 1$ . Because the sample is of finite size, the Lévy walk is truncated at  $t = d_{\max}/v$ , where  $d_{\max}$  is the maximum step length, determined by the sample thickness.



Real physical samples are intrinsically of finite size, which means that the largest step size of the Lévy flight is limited by the sample size. This introduces a cutoff in the step-length distribution and results in a so-called truncated Lévy flight. On length scales greater than this cutoff, the transport is expected to recover normal diffusive behaviour<sup>29</sup>. We investigated this by running a series of Monte Carlo simulations in which we studied a random walk in a two-dimensional system similar to our samples, namely a scattering medium where disk-shaped regions are introduced without scattering elements. The diameter distribution of these two-dimensional disks was chosen, following the same reasoning as above, as  $P_s(d) = 1/d^2$ . We simulated the evolution with time of the averaged squared displacement of light propagating in this system. The results of these simulations (Fig. 5) show superdiffusive behaviour that, on a very long timescale, develops into normal diffusive behaviour. The parameter  $\gamma$  of the superdiffusive expansion was found to be close to two, as expected for a Lorentzian Lévy flight. The timescale of the transition from superdiffusive to diffusive transport is given by the time necessary to probe all possible step lengths:  $t_{\text{trans}} = d_{\text{max}}/v$ , where  $d_{\text{max}}$  is the greatest step length and  $v$  is the velocity of the random walker. In our samples, the thickness was equal to this cutoff length (greatest sphere diameter). As a result, the effect of the cutoff can be expected to be negligible within the signal-to-noise ratio of our experiment.

We have shown that it is possible to make disordered optical materials with controllable step-length distributions. In particular, we have made superdiffusive optical materials permitting optical Lévy flights. The physics of light transport is closely related to the transport of electrons and matter waves, and important analogies like the optical Hall effect, weak and strong localization of light, and correlations in laser speckle have been identified in recent years. The question of how these phenomena are manifest in Lévy glass is still completely open. The procedure that we have used to synthesize Lévy glass is reproducible and can be implemented on a large (industrial) scale. Our techniques could be used in the development of new opaque optical materials, such as paints with particular visual effects and lasers based on superdiffusive feedback.

Received 19 September 2007; accepted 26 March 2008.

1. Kubo, R., Toda, M. & Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, Berlin, 1991).
2. Mandelbrot, B. *The Fractal Geometry of Nature* (Freeman, New York, 1977).
3. Lévy, P. *Théorie de l'Addition des Variables Aléatoires* (Gauthier-Villars, Paris, 1954).
4. Bartumeus, F., Da Luz, M. G. E., Viswanathan, G. M. & Catalan, J. Animal search strategies: A quantitative random-walk analysis. *Ecology* **86**, 3078–3087 (2005).
5. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
6. Corral, A. Universal earthquake-occurrence jumps, correlations with time, and anomalous diffusion. *Phys. Rev. Lett.* **97**, 178501 (2006).

7. Geisel, T., Nierwetberg, J. & Zacherel, A. Accelerated diffusion in Josephson junctions and related chaotic systems. *Phys. Rev. Lett.* **54**, 616–619 (1985).
8. Drysdale, P. M. & Robinson, P. A. Lévy random walks in finite systems. *Phys. Rev. Lett.* **58**, 5382–5394 (1998).
9. Frenkel, D., van Luijn, F. & Binder, P. Evidence for universal asymptotic decay of velocity fluctuations in Lorentz gases. *Europhys. Lett.* **20**, 7–12 (1992).
10. Ott, A., Bouchaud, J., Langevin, D. & Urbach, W. Anomalous diffusion in living polymers: a genuine Lévy flight? *Phys. Rev. Lett.* **65**, 2201–2204 (1990).
11. Solomon, T. H., Weeks, E. R. & Swinney, H. L. Observation of anomalous diffusion and Lévy flights in a two-dimensional rotating flow. *Phys. Rev. Lett.* **71**, 3975–3978 (1993).
12. Sheng, P. *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego, 1995).
13. Rikken, G. L. J. A. & van Tiggelen, B. A. Observation of magnetically induced transverse diffusion of light. *Nature* **381**, 54–55 (1996).
14. Sapienza, R. et al. The optical analogue of electronic Bloch oscillations. *Phys. Rev. Lett.* **91**, 263902 (2003).
15. Scheffold, F. & Maret, G. Universal conductance fluctuations of light. *Phys. Rev. Lett.* **81**, 5800–5803 (1998).
16. Yodh, A. G., Chance, B., Boas, D. A. & O'Leary, M. Object imaging using diffuse light. US patent 5,917 190 (1999).
17. Cao, H. et al. Random laser action in semiconductor powder. *Phys. Rev. Lett.* **82**, 2278–2281 (1999).
18. Ripoll, J. & Ntziachristos, V. From finite to infinite volumes: removal of boundaries in diffuse wave imaging. *Phys. Rev. Lett.* **96**, 173903 (2006).
19. Shlesinger, M. F., Zaslavsky, G. M. & Klafter, J. Strange kinetics. *Nature* **363**, 31–37 (1993).
20. Shlesinger, M. F. Physics in the noise. *Nature* **411**, 641 (2001).
21. Mandelbrot, B. The variation of certain speculative prices. *J. Bus.* **36**, 394–419 (1963).
22. Mantegna, R. & Stanley, H. Scaling behaviour of an economic index. *Nature* **376**, 46–49 (1995).
23. Lepri, S., Cavalieri, S., Oppo, G.-L. & Wiersma, D. S. Statistical regimes of random laser fluctuations. *Phys. Rev. A* **75**, 063820 (2007).
24. Sharma, D., Ramachandran, H. & Kumar, N. Lévy statistical fluctuations from a random amplifying medium. *Fluct. Noise Lett.* **6**, 95–101 (2006).
25. Bardou, F., Bouchaud, J. P., Emile, O., Aspect, A. & Cohen-Tannoudji, C. Subrecoil laser cooling and Lévy flights. *Phys. Rev. Lett.* **72**, 203–206 (1994).
26. Nolan, J. P. *Stable Distributions* (Birkhäuser, Boston, 2002).
27. Metzler, R., Chechkin, A., Gonchar, V. & Klafter, J. Some fundamental aspects of Lévy flights. *Chaos Solitons Fractals* **34**, 129–142 (2007).
28. Buldyrev, S. V. et al. Average time spent by Lévy flights and walks on an interval with absorbing boundaries. *Phys. Rev. E* **64**, 041108 (2001).
29. Mantegna, R. N. & Stanley, H. E. Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. *Phys. Rev. Lett.* **73**, 2946–2949 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We wish to thank A. Lagendijk for discussions and for reading the manuscript. Also we thank R. Righini and M. Colocci for continuous support, the entire Optics of Complex Systems group at LENS for discussions. This project has been financed by the ATLAS program of the European Commission, as well as the European Network of Excellence PHOREMOST.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.S.W. ([wiersma@lens.unifi.it](mailto:wiersma@lens.unifi.it)).

# Gelation of particles with short-range attraction

Peter J. Lu<sup>1</sup>, Emanuela Zaccarelli<sup>3,4</sup>, Fabio Ciulla<sup>3</sup>, Andrew B. Schofield<sup>5</sup>, Francesco Sciortino<sup>3,4</sup> & David A. Weitz<sup>1,2</sup>

Nanoscale or colloidal particles are important in many realms of science and technology. They can dramatically change the properties of materials, imparting solid-like behaviour to a wide variety of complex fluids<sup>1,2</sup>. This behaviour arises when particles aggregate to form mesoscopic clusters and networks. The essential component leading to aggregation is an interparticle attraction, which can be generated by many physical and chemical mechanisms. In the limit of irreversible aggregation, infinitely strong interparticle bonds lead to diffusion-limited cluster aggregation<sup>3</sup> (DLCA). This is understood as a purely kinetic phenomenon that can form solid-like gels at arbitrarily low particle volume fraction<sup>4,5</sup>. Far more important technologically are systems with weaker attractions, where gel formation requires higher volume fractions. Numerous scenarios for gelation have been proposed, including DLCA<sup>6</sup>, kinetic or dynamic arrest<sup>4,7–10</sup>, phase separation<sup>5,6,11–16</sup>, percolation<sup>4,12,17,18</sup> and jamming<sup>8</sup>. No consensus has emerged and, despite its ubiquity and significance, gelation is far from understood—even the location of the gelation phase boundary is not agreed on<sup>5</sup>. Here we report experiments showing that gelation of spherical particles with isotropic, short-range attractions is initiated by spinodal decomposition; this thermodynamic instability triggers the formation of density fluctuations, leading to spanning clusters that dynamically arrest to create a gel. This simple picture of gelation does not depend on microscopic system-specific details, and should thus apply broadly to any particle system with short-range attractions. Our results suggest that gelation—often considered a purely kinetic phenomenon<sup>4,8–10</sup>—is in fact a direct consequence of equilibrium liquid–gas phase separation<sup>5,13–15</sup>. Without exception, we observe gelation in all of our samples predicted by theory and simulation to phase-separate; this suggests that it is phase separation, not percolation<sup>12</sup>, that corresponds to gelation in models for attractive spheres.

Gelation occurs in a wide range of systems where particles attract each other<sup>2,5–8,11,12,15–18</sup>. When this attraction is infinitely strong, particles form permanent bonds and grow as fractal clusters that, in turn, bond irreversibly, and can ultimately span the system as a solid-like gel, even as particle volume fraction  $\phi$  tends to zero (refs 4, 5, 12, 19). This DLCA limit occurs in many colloidal systems where the interparticle attraction strength,  $U$ , is much larger than the thermal energy  $k_B T$  (refs 4, 5, 12); examples include gold<sup>3,20</sup>, silica<sup>3</sup>, polymeric lattices<sup>3,6,19</sup>, calcium carbonate<sup>21</sup>, alumina<sup>2</sup> and silicon carbide<sup>2</sup>. Because bonds once formed never break, DLCA is governed entirely by diffusion; it has thus been considered a purely kinetic phenomenon<sup>3</sup>. Other mechanisms can cause kinetic arrest at far higher  $\phi$  (ref. 5). Above  $\phi \approx 0.58$ , particles can arrest because of crowding to form repulsive glasses, even when  $U = 0$ ; weakly attractive particles can form attractive glasses at lower  $\phi$  (ref. 5). Because glasses and DLCA are observed in the same experimental systems, they have been linked within unified pictures of kinetic arrest<sup>4,7,9,10</sup> or jamming<sup>8</sup>.

More generally, the onset of gelation can be parameterized by three quantities, namely  $\phi$ ,  $U/k_B T$  and  $\xi$ . The last is the range of

the attractive potential in units of  $a$ , the particle radius<sup>4,22</sup>. These three parameters define a three-dimensional state diagram in which a gelation surface demarcates the well-defined boundary between liquid-like and solid-like behaviour. Many important attraction mechanisms that drive gelation are short-range ( $\xi < 0.1$ ), including van der Waals forces<sup>8,16,21</sup>, surface chemistry<sup>2,17,18</sup>, hydrophobic effects<sup>7</sup> and some depletion interactions<sup>9,15,23</sup>. Numerous explanations have been advanced for gelation in this small- $\xi$  limit to predict the fluid–solid boundary in the  $U$ – $\phi$  plane. Non-equilibrium, kinetics-based models have extended the DLCA model to lower  $U/k_B T$  by treating bond breakage probabilistically<sup>6,12,20</sup>; have connected the gelation boundary to the percolation threshold<sup>4,12,17,18</sup>; and have extended the glass transition to lower  $\phi$  with mode-coupling theory applied to local arrest of individual particles<sup>9</sup>, to arrest of clusters<sup>4</sup>, and in concert with microscopic modelling of the interparticle attractive potential<sup>23</sup>. Thermodynamic models consider gelation initiated by fluid–crystal<sup>11</sup>, liquid–gas<sup>6,14,15</sup>, or polymer-like ‘viscoelastic’<sup>16</sup> phase separation, which may arrest owing to percolation<sup>12</sup> or a glass transition<sup>4</sup>. These models make strikingly disparate predictions: there is no agreement on either the gelation mechanism, or the location of the gelation boundary<sup>5,12,23</sup>.

Here we explore gelation experimentally with a widely-used model colloid–polymer system<sup>6,11,22</sup>, where  $U/k_B T$  and  $\xi$  are controlled by the polymer size and free-volume concentration  $c_p$ , but in a fashion that is not precisely known. Fixing  $\phi = 0.045 \pm 0.005$  and  $\xi = 0.059$ , we mix samples at various  $c_p$ ; we summarize the samples studied by plotting their values of  $c_p$ , normalized by the polymer overlap concentration  $c_p^*$ , in the phase diagrams shown in Fig. 1a, b. We eliminate gravitational sedimentation on multiple-day timescales by meticulously matching the colloid and solvent densities to within  $< 10^{-4}$ . After breaking up particle aggregates by shearing, we observe sample evolution with a high-speed confocal microscope<sup>24</sup>.

We observe two phases. In samples with low  $c_p$ , below the experimental gelation boundary  $c_p^g$ , we observe a fluid of many clusters that is stable for days; we show a full three-dimensional image of these clusters in the fluid phase for a sample with  $c_p = 3.20 \pm 0.03 \text{ mg ml}^{-1}$ , the closest fluid-phase value below  $c_p^g$ , in Fig. 1c and in Supplementary Video 1. By contrast, in samples with  $c_p > c_p^g$ , particles aggregate immediately into clusters, which in turn form a network that spans the macroscopic sample. This network subsequently arrests to create a gel, which we illustrate for a sample with  $c_p = 3.31 \pm 0.03 \text{ mg ml}^{-1}$ , the closest gel-phase value above  $c_p^g$ , in Fig. 1d and in Supplementary Video 2. The gel undergoes no major structural rearrangement for days, even though it exchanges particles with a dilute gas, shown in Supplementary Video 3. These phases are separated by a very sharp boundary: the gel and fluid illustrated differ in  $c_p$  by only a few per cent. Our observation of only these two dramatically different phases contrasts findings of more complex phase behaviour in non-buoyancy-matched systems, where sedimentation can shift or obscure the observed phase boundaries<sup>6,9,12,15,21</sup>.

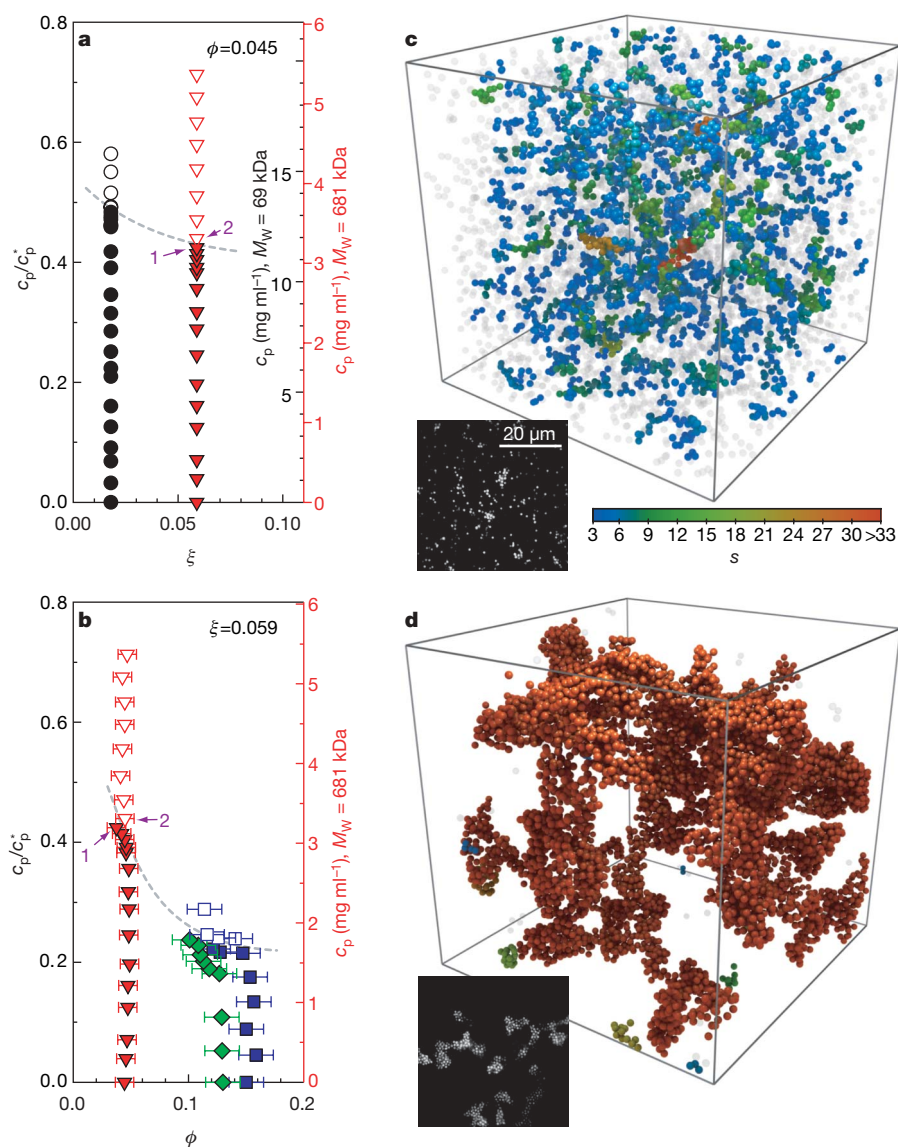
<sup>1</sup>Department of Physics, <sup>2</sup>SEAS, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Dipartimento di Fisica, <sup>4</sup>CNR-INFM-SOFT, Università di Roma La Sapienza, Piazzale A. Moro 2, 00185 Roma, Italy. <sup>5</sup>The School of Physics, University of Edinburgh, Edinburgh EH9 3JZ, UK.

Locating the gelation boundary in general requires a means to compare among experiments and with theory or simulation, using universal thermodynamic quantities, like  $U/k_B T$ , instead of system-specific parameters, like  $c_p$  (refs 9, 23). Unfortunately, it is impossible to precisely determine  $U/k_B T$  from a known  $c_p$ , even using microscopic models for the potential. Instead, we use the finding that the behaviour of an attractive particle system for  $\xi < 0.1$  depends not on the shape of the potential, but only on its reduced second virial coefficient,  $B_2^* \equiv (3/8a^3) \int_0^\infty (1 - \exp(-U(r)/k_B T)) r^2 dr$  (ref. 25).

After each fluid sample has reached its long-term steady state, we determine its cluster mass distribution  $n(s)$ , the fraction of total clusters that contain  $s$  particles. We then simulate hard spheres with isotropic short-range attractions at the same  $\phi$ , determining  $n(s)$  for different values of  $B_2^*$ . For each experimental  $n(s)$ , we select the

closest-matching simulated  $n(s)$  using a least-squares minimization. This allows us to associate each  $c_p$  with a unique  $B_2^*$ , with no adjustable parameters. These fits all work remarkably well, irrespective of the interparticle attractive potential shape, so long as the potentials have the same  $B_2^*$ , as shown in Fig. 2. Identical  $n(s)$  are observed for the square-well, generalized Lennard–Jones, and Asakura–Oosawa forms, commonly used for colloid–polymer mixtures<sup>9,23,26</sup>, substantiating our  $c_p$ – $B_2^*$  mapping even though the exact experimental potential shape remains unknown. Measuring  $n(s)$  requires only straightforward counting of particle bonds; by contrast, determining  $B_2^*$  with similar precision from scattering<sup>18</sup> or radial distribution functions<sup>27</sup> requires far more accurate identification of particle positions.

From  $B_2^*$ , other thermodynamic quantities can be derived directly, including  $k_B T/U$  for different potential forms<sup>25</sup>. Considerable insight



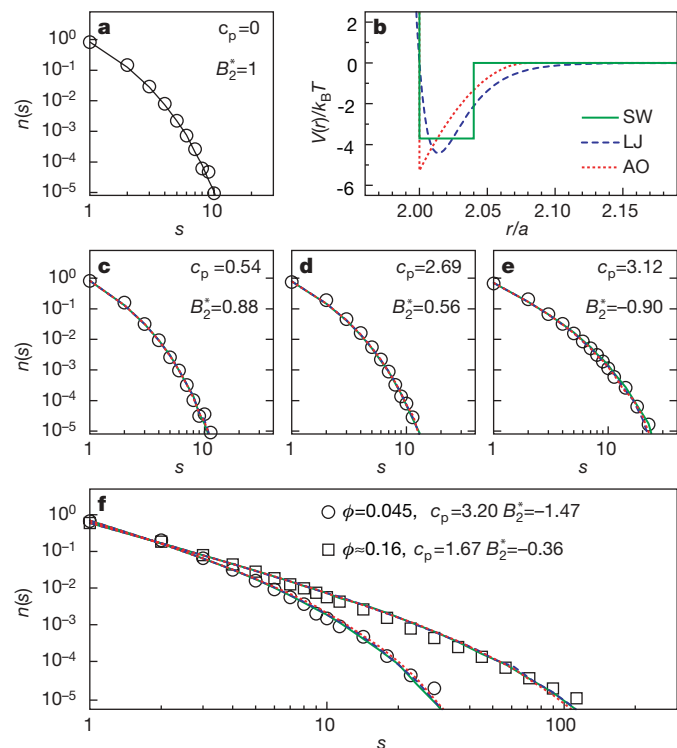
**Figure 1 | Composition and structure of experimental gel and fluid samples.** **a**, Experimental samples in a  $c_p/c_p^*$  and  $\xi$  phase diagram for constant  $\phi = 0.045$ . Black circles and red triangles indicate samples with 69 kDa and 681 kDa polystyrene polymers, respectively. Solid symbols mark fluid samples; open symbols, gels. Actual measured  $c_p$  values are on secondary vertical axes of the same colour at right. **b**, Experimental samples in a  $c_p/c_p^*$  and  $\phi$  phase diagram for constant  $\xi = 0.059$ , with  $c_p$  of the 681 kDa polymer used in all samples indicated on the secondary red axis at right. Error bars represent the variation in  $\phi$  for different particle configurations from the same sample. In **a** and **b**, dashed grey gelation boundaries are drawn to guide

the eye. **c**, 3D reconstruction ( $56 \times 56 \times 56 \mu\text{m}^3$ ), and (inset) single 2D confocal microscope image, for the fluid with the highest  $c_p = 3.20 \text{ mg ml}^{-1}$ . The fluid's clusters are coloured by their mass  $s$  (number of particles) according to the colour bar, with monomers and dimers rendered in transparent grey to improve visibility. **d**, Reconstruction and confocal image of the gel with the lowest  $c_p = 3.31 \text{ mg ml}^{-1}$  shown at same scale, containing a single spanning cluster. Samples in **c** and **d** are in the long-time steady state four hours after mixing; their compositions are marked in **a** and **b** with the purple numerals 1 and 2, respectively.



is obtained by using  $n(s)$  fits to determine the values of  $k_B T/U$ , calculated for an Asakura–Oosawa potential with  $\xi = 0.059$  to match the experiment, and plotting these as a function of  $c_p$  for all fluid samples. The data exhibit an unexpected linear dependence near the experimentally determined gelation boundary at  $c_p^g = 3.25 \pm 0.05 \text{ mg ml}^{-1}$ , as shown in Fig. 3a. We calculate the onset of phase separation both in the Baxter model and with simulation, which, in all cases, yield identical results. Remarkably, these correspond precisely to the experimentally determined value of  $k_B T/U$  at the gel boundary, as shown in Fig. 3a. This suggests that the gel boundary occurs exactly at the boundary of phase separation. Because the spinodal and binodal lines are very close for all short-range potentials, such as those here, we do not observe nucleation and growth—instead, the observed phase separation is always driven by spinodal decomposition.

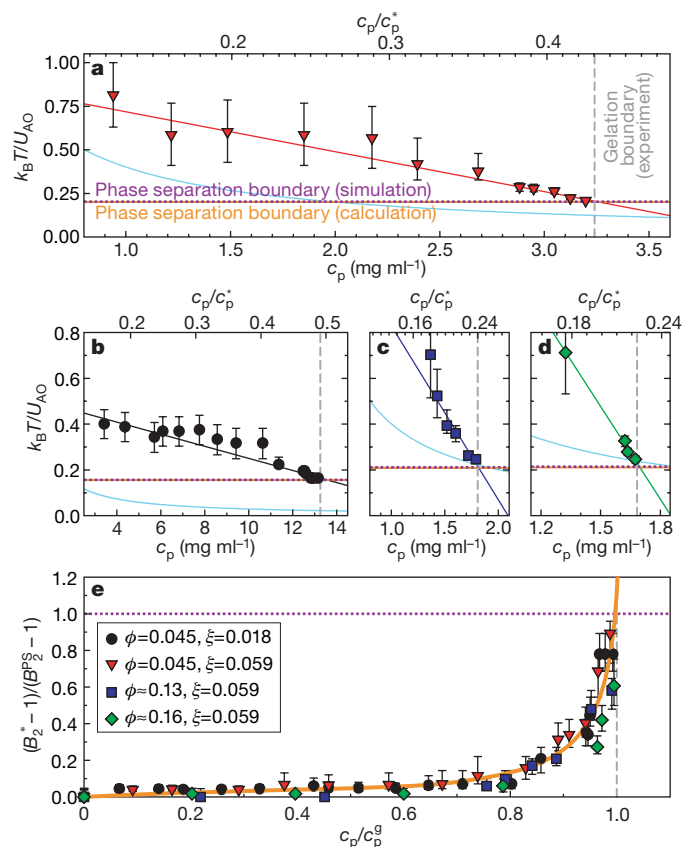
To confirm the generality of these results, we repeat the experiment for different  $\phi$  and  $\xi$ . Again fixing  $\xi = 0.059$ , we create additional samples at  $\phi \approx 0.13$  and  $\phi \approx 0.16$ , as shown in the phase diagram in Fig. 1b. Increasing  $\phi$  results in larger clusters, whose mass distribution broadens to more closely resemble a power law, as shown in Fig. 2f; this is reminiscent of an approach to the critical



**Figure 2 | Comparisons among cluster mass distributions  $n(s)$  for  $\xi = 0.059$ .** **a**, Comparison at  $\phi = 0.045$  between experimental data for  $c_p = 0$  (circles), and simulation results for a hard-sphere potential ( $U/k_B T = 0$  and  $B_2^* = 1$ ; solid line), demonstrating an exact match. In this and all panels, value of  $c_p$  is in  $\text{mg ml}^{-1}$ . **b**, Three potentials at the same  $B_2^*$  used to generate  $n(s)$  in simulations with finite attractions; solid green, dashed blue and dotted red lines denote square-well (SW), generalized Lennard–Jones (LJ) and Asakura–Oosawa (AO) potentials, respectively. Example potentials shown for  $B_2^* = -1.47$ . **c–e**, Example comparisons at  $\phi = 0.045$  between experimental  $n(s)$ , marked by circles, and simulation  $n(s)$ , by lines coloured as the corresponding potentials in **b**. **c**,  $c_p = 0.54 \text{ mg ml}^{-1}$  and  $B_2^* = 0.88$ . **d**,  $c_p = 2.69 \text{ mg ml}^{-1}$  and  $B_2^* = 0.56$ . **e**,  $c_p = 3.12 \text{ mg ml}^{-1}$  and  $B_2^* = -0.90$ . **f**, Comparisons for the fluids with the highest  $c_p$  closest to the gel boundary at  $c_p^g$ . Circles denote the fluid with  $\phi = 0.045$  ( $c_p = 3.20 \text{ mg ml}^{-1}$  and  $B_2^* = -1.47$ ; sample illustrated in Fig. 1c). Squares denote the fluid with  $\phi \approx 0.16$  ( $c_p = 1.67 \text{ mg ml}^{-1}$  and  $B_2^* = -0.36$ ), whose significantly larger clusters are expected as the  $\phi_c = 0.27$  critical point is approached. All data sets match exactly, confirming that  $n(s)$  both usefully maps experimental to simulation results and does not depend on potential shape.

point predicted at  $\phi_c \approx 0.27$  (ref. 28). In addition, for  $\phi = 0.045$ , we also reduce  $\xi$  to 0.018; this yields more tenuous, branched, thinner clusters<sup>22</sup>. These samples are shown in the phase diagram in Fig. 1a. In all cases, the experimentally determined value of  $k_B T/U$  at the gelation boundary coincides exactly with the theoretical phase separation boundary, as shown in Fig. 3b–d. Finally, we consider the dependence of  $B_2^* - 1$ , normalized by the value at the phase separation boundary, as a function of  $c_p/c_p^g$ . Unexpectedly, despite significant variation in cluster morphology, all sample data scale onto a single master curve, shown in Fig. 3e. This highlights the similarities in behaviour of all samples on approach to the spinodal line and points to a universal mechanism for gelation.

These data suggest that, for isotropic short-range interactions, all gelation is triggered by spinodal decomposition, a phase separation process driven by a thermodynamic instability. If this is so, then we should independently observe other characteristics of equilibrium phase separation in samples that form gels. One such feature is the coexistence of gel and colloidal gas: we observe occasional exchange of particles between gas and gel, as shown in Supplementary Video 3;



**Figure 3 | Comparison of  $n(s)$  mapping of experimental  $c_p$  to  $k_B T/U$ .** Data are shown for **a**,  $\phi = 0.045$  and  $\xi = 0.059$ , **b**,  $\phi = 0.045$  and  $\xi = 0.018$ , **c**,  $\phi \approx 0.13$  and  $\xi = 0.059$ , and **d**,  $\phi \approx 0.16$  and  $\xi = 0.059$ . Grey dashed vertical lines demarcate the experimental gelation boundary at  $c_p^g$ ; horizontal lines demarcate the theoretical phase separation boundary calculated in the Baxter model (orange solid line) and with simulation (purple dotted line), which always coincide. Coloured symbols (as used in Fig. 1a, b and shown in the key in **e**) with best-fit lines represent the results of the  $n(s)$  mapping illustrated in Fig. 2; error bars correspond to the uncertainty from the least-squares fitting. The experimental gelation boundary exactly matches the theoretical phase separation boundary for all  $\phi$  and  $\xi$ ; by contrast, analytic approximation to the Asakura–Oosawa potential, shown in light blue, does not match at all. **e**, Mapping between  $c_p$  and  $B_2^* - 1$  for all fluid samples, where  $c_p$  is normalized by  $c_p^g$  (grey dashed vertical line), and  $B_2^* - 1$  by  $B_2^{*g} - 1$ , its value at the phase separation boundary (purple dotted horizontal line). All data collapse onto a single master curve, highlighted with an orange line to guide the eye. Gelation exhibits universal scaling independent of  $\phi$ ,  $\xi$  or shape of the short-range potential.

this is not readily explained by kinetic gelation models based on local arrest<sup>9,10</sup>. An even more distinctive hallmark of spinodal decomposition is the development of a peak in the static structure factor  $S(q)$  at finite scattering vector  $q$  (refs 19, 29). We again observe this: in fluid samples with  $\phi = 0.045$ ,  $\xi = 0.059$  and  $c_p < c_p^g$ ,  $S(q)$  shows only a slight rise at low  $q$ ; however, increasing  $c_p$  by just a few per cent across  $c_p^g$  increases the height of the peak in  $S(q)$  by two orders of magnitude, as shown in Fig. 4a. Further distinguishing characteristics of spinodal decomposition occur in the temporal evolution of  $S(q)$ , where the peak narrows and moves towards lower  $q$ , and in its first moment  $q_1(t)$ , which exhibits a power law dependence. Once again, the gel samples unambiguously demonstrate these features: at the earliest times, the peak in  $S(q)$  narrows and moves to lower  $q$ , as shown in Fig. 4b; moreover,  $q_1(t)$  scales as  $t^{-1/6}$ , as shown in Fig. 4c, exactly as in molecular spinodal decomposition<sup>30</sup>. Two hours after mixing, the spinodal decomposition towards the equilibrium phase-separated

state is interrupted, as the sample dynamically arrests to form a gel;  $S(q)$  and  $q_1$  no longer change with time, as shown in Fig. 4b–c. Similar dynamics for  $S(q)$  are observed in all gel samples, further demonstrating that liquid–gas spinodal decomposition ubiquitously induces gelation for short-range potentials.

Together, these results provide strong, quantitative physical evidence that the gelation boundary for short-range attractive particles is precisely equivalent to the boundary for equilibrium liquid–gas phase separation. Gelation requires spinodal decomposition to generate the clusters that span the system and dynamically arrest. Our findings experimentally confirm previous theoretical predictions<sup>5,13,14</sup>, and support the suggestion that the ostensibly purely kinetic DLCA regime is in fact a deeply quenched limit of spinodal decomposition<sup>19,29</sup>. Thus, thermodynamic instability appears to drive all gelation of particles with isotropic short-range attractions.

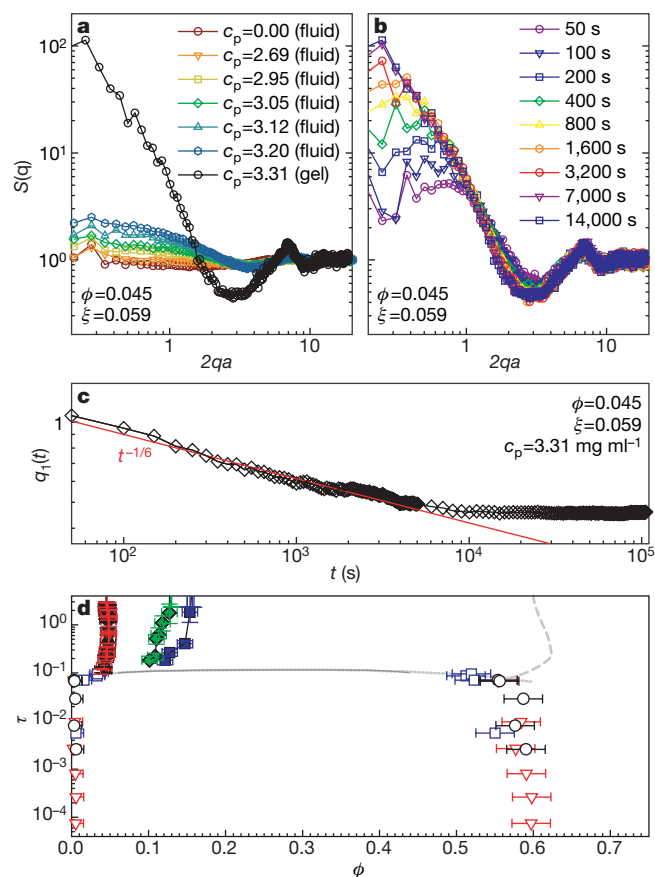
We cannot harmonize our results with predictions from phase separation that is not liquid–gas<sup>11,16</sup>, nor from purely kinetic paradigms<sup>4,8–10</sup>. However, the expression of these predictions as system-specific  $c_p/c_p^*$  values calculated for the Asakura–Oosawa potential may affect comparison of results. To test this, we plot  $k_B T/U$  versus  $c_p/c_p^*$  for an analytic approximation to the Asakura–Oosawa potential<sup>9</sup> in Fig. 3a–d, which in all cases dramatically misses the actual potential strength determined from the  $n(s)$  mapping; this corroborates previous findings that the Asakura–Oosawa model does not quantitatively describe colloid–polymer mixtures<sup>23,26,27</sup>.

Instead, universal system-independent parameters, such as  $B_2^*$  (refs 5, 12, 13, 15, 17, 18) and  $\phi$ , allow meaningfully quantitative comparison between different experiments and with theory. We present such a comparison, as a universal phase diagram for short-range gelation, in Fig. 4d. Without exception, all samples predicted within the Baxter model to phase-separate form gels. This suggests that the gelation line coincides with the phase separation boundary in the Baxter model; other isotropic short-range potentials have similar behaviour. For gel samples, we estimate the volume fractions in both colloidal gas and gel phases by numerically determining the free volume accessible to a test particle of radius  $a$ ; we consider this the total volume of the gas phase, and assign the remaining volume to the gel. Surprisingly, we find that all spanning gel clusters have  $\phi_g \approx 0.55$ , independent of both  $c_p$  and the average  $\phi$  before phase separation. We never observe arrested spanning clusters with significantly lower  $\phi_g$ ; the attractive glass line must therefore intersect the phase separation boundary at  $\phi \approx 0.55$  (refs 5, 13), consistent with the origin of kinetic arrest arising from the dense phase undergoing an attractive glass transition<sup>5,13</sup>. Furthermore,  $\phi_g$  does not decrease with increasing attraction strength<sup>4,7,9</sup>, suggesting that the attractive glass line does not extend into the phase separation region, but instead follows its boundary.

Our results could shed light on non-equilibrium behaviour in technological systems. Even approximate measures of structural parameters, such as  $n(s)$ , may, when compared with simulations, allow mapping between thermodynamic quantities and experimental parameters when even the rough form of the potential cannot be measured. Moreover, because the onset of non-equilibrium behaviour is in fact governed by equilibrium phase separation, thermodynamic calculations may facilitate quantitative prediction of product stability, a critically important problem in the formulation and manufacture of commercial complex fluids.

## METHODS SUMMARY

We suspend polymethylmethacrylate (PMMA) colloidal spheres of radius  $a = 560$  nm in a solvent mixture with matching buoyancy and refractive index, adding an organic salt to screen Coulombic repulsion and linear polystyrene to induce a depletion attraction<sup>22,24</sup>. We determine the radii of colloid and polymer coils with light scattering. We image all samples in a high-speed, automated confocal microscope<sup>24</sup>, collecting 181 images at 10 frames per second in each three-dimensional (3D) stack, which occupies a  $60 \times 60 \times 60 \mu\text{m}^3$  cube within the sample. We use previously described image-processing software<sup>24</sup> to determine the 3D positions of all colloidal particles in each sample. In total, we collected half a terabyte of image data and located  $\sim 10^8$  particles. We use



**Figure 4 | Spinodal decomposition in samples that form gels.** **a**,  $S(q)$  in the long-time steady-state limit for fluid samples at  $\phi = 0.045$  and  $\xi = 0.059$  with  $c_p \leq 3.20$  mg ml<sup>-1</sup> (coloured symbols) and the gel sample with  $c_p = 3.31$  mg ml<sup>-1</sup> (black circles). Blue hexagons and black circles denote the fluid and gel samples illustrated in Fig. 1c and d, respectively. All fluid samples show  $S(q)$  rising slightly at low  $q$  as  $c_p \rightarrow c_p^g$ . As  $c_p$  crosses  $c_p^g$  into the gel region,  $S(q)$  develops a significant peak two orders of magnitude higher. **b**, Time evolution of  $S(q)$  for this gel. Immediately after sample homogenization, a finite- $q$  peak grows, narrows, and shifts to lower  $q$ , as expected for spinodal decomposition. **c**,  $q_1(t)$  (black diamonds) follows a  $t^{-1/6}$  power law (red line), another hallmark of spinodal decomposition. After two hours, the sample arrests to form a gel, and  $S(q)$  and  $q_1$  do not change. **d**, Universal phase diagram of the Baxter parameter  $\tau \equiv 1/4(B_2^* - 1)$  and  $\phi$  for all samples, with symbols as in Fig. 1a, b and estimates of  $\phi$  shown for both gas and gel phases after phase separation. Error bars represent the variation in  $\phi$  for different particle configurations from the same sample. All samples predicted to phase-separate within the Baxter model, falling below the theoretical phase separation boundary from ref. 28 (solid grey line), form gels with the same  $\phi_g$ . Speculative extensions of this boundary (dotted grey line) and of the glass transition (dashed grey line) are plotted to guide the eye.

Pixar's RenderMan (<https://renderman.pixar.com>) to create 3D reconstructions. We perform simulations of fluid samples of 10,000 particles in a cubic box with periodic boundary conditions for several values of  $B_2^*$ , using several simulated potentials: a hard-sphere potential, a square-well potential of width  $0.04a$ , an Asakura–Oosawa potential of maximum width  $0.08a$ , and a generalized  $2\alpha$ - $\alpha$  Lennard–Jones potential with exponent  $\alpha = 100$ . Following a constant-temperature equilibration run, we generate 100 independent realizations in the micro-canonical ensemble for subsequent analysis. We estimate the spinodal line following the temperature-dependence of the energy and of the small angle structure factor within simulations<sup>13</sup>, and using the energy route in the Percus–Yevick approximation to the Baxter model for hard spheres with an infinitesimally short attraction range<sup>28</sup>. We use the same procedure in experiment and simulation to assign particles to clusters by considering which particles share common bonds; two particles are considered bonded if they are separated by less than the bond distance  $r_b$ , fixed by matching the  $c_p = 0$  cluster-mass distributions. We use a least-squares minimization to best match numerical distributions to the experimental results with no free parameters.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 11 December 2007; accepted 14 March 2008.**

- Mezzenga, R., Schurtenberger, P., Burbidge, A. & Michel, M. Understanding foods as soft materials. *Nature Mater.* **4**, 729–740 (2005).
- Gauckler, L. J., Graule, Th. & Baader, F. Ceramic forming using enzyme catalyzed reactions. *Mater. Chem. Phys.* **61**, 78–102 (1999).
- Lin, M. Y. et al. Universality in colloid aggregation. *Nature* **339**, 360–362 (1989).
- Kroy, K., Cates, M. E. & Poon, W. C. K. Cluster mode-coupling approach to weak gelation in attractive colloids. *Phys. Rev. Lett.* **92**, 148302 (2004).
- Zaccarelli, E. Colloidal gels: Equilibrium and non-equilibrium routes. *J. Phys. Condens. Matter* **19**, 323101 (2007).
- de Hoog, E. H. A., Kegel, W. K., van Blaaderen, A. & Lekkerkerker, H. N. W. Direct observation of crystallization and aggregation in a phase-separating colloid-polymer suspension. *Phys. Rev. E* **64**, 021407 (2001).
- Cardinaux, F., Gibaud, T., Stradner, A. & Schurtenberger, P. Interplay between spinodal decomposition and glass formation in proteins exhibiting short-range attractions. *Phys. Rev. Lett.* **99**, 118301 (2007).
- Trappe, V., Prasad, V., Cipelletti, L., Segre, P. N. & Weitz, D. A. Jamming phase diagram for attractive particles. *Nature* **411**, 772–775 (2001).
- Bergenholtz, J., Poon, W. C. K. & Fuchs, M. Gelation in model colloid-polymer mixtures. *Langmuir* **19**, 4493–4503 (2003).
- Puertas, A. M., Fuchs, M. & Cates, M. E. Competition between glass transition and liquid-gas phase separation in attracting colloids. *J. Phys. Condens. Matter* **19**, 205140 (2007).
- Pusey, P. N., Pirie, A. D. & Poon, W. C. K. Dynamics of colloid-polymer mixtures. *Physica A* **201**, 322–331 (1993).
- Poon, W. C. K. & Haw, M. D. Mesoscopic structure formation in colloidal aggregation and gelation. *Adv. Colloid Interface Sci.* **73**, 71–126 (1997).
- Foffi, G., De Michele, C., Sciortino, F. & Tartaglia, P. Scaling of dynamics with the range of interaction in short-range attractive colloids. *Phys. Rev. Lett.* **94**, 078301 (2005).
- Charbonneau, P. & Reichman, D. R. Systematic characterization of thermodynamic and dynamical phase behavior in systems with short-ranged attraction. *Phys. Rev. E* **75**, 011507 (2007).
- Buzzaccaro, S., Rusconi, R. & Piazza, R. "Sticky" hard spheres: Equation of state, phase diagram, and metastable gels. *Phys. Rev. Lett.* **99**, 098301 (2007).
- Tanaka, H., Nishikawa, Y. & Koyama, T. Network-forming phase separation of colloidal suspensions. *J. Phys. Condens. Matter* **17**, L143–L153 (2005).
- Verduin, H. & Dhont, J. K. G. Phase diagram of a model adhesive hard-sphere dispersion. *J. Colloid Interface Sci.* **172**, 425–437 (1995).
- Grant, M. C. & Russel, W. B. Volume-fraction dependence of elastic moduli and transition temperatures for colloidal silica gels. *Phys. Rev. E* **47**, 2606–2614 (1993).
- Carpinetti, M. & Giglio, M. Spinodal-type dynamics in fractal aggregation of colloidal clusters. *Phys. Rev. Lett.* **68**, 3327–3330 (1992).
- Liu, J., Shih, W. Y., Sarikaya, M. & Aksay, I. A. Fractal colloidal aggregates with finite interparticle interactions: Energy dependence of the fractal dimension. *Phys. Rev. A* **41**, 3206–3213 (1990).
- Allain, C., Cloitre, M. & Wafra, M. Aggregation and sedimentation in colloidal suspensions. *Phys. Rev. Lett.* **74**, 1478–1481 (1995).
- Lu, P. J., Conrad, J. C., Wyss, H. M., Schofield, A. B. & Weitz, D. A. Fluids of clusters in attractive colloids. *Phys. Rev. Lett.* **96**, 028306 (2006).
- Shah, S. A., Chen, Y. L., Schweizer, K. S. & Zukoski, C. F. Phase behavior and concentration fluctuations in suspensions of hard spheres and nearly ideal polymers. *J. Chem. Phys.* **118**, 3350–3361 (2003).
- Lu, P. J., Sims, P. A., Oki, H., Macarthur, J. B. & Weitz, D. A. Target-locking acquisition with real-time confocal (TARC) microscopy. *Opt. Express* **15**, 8702–8712 (2007).
- Noro, M. G. & Frenkel, D. Extended corresponding-states behavior for particles with variable range attractions. *J. Chem. Phys.* **113**, 2941–2944 (2000).
- Bolhuis, P. G., Louis, A. A. & Hansen, J.-P. Influence of polymer-excluded volume on the phase-behavior of colloid-polymer mixtures. *Phys. Rev. Lett.* **89**, 128302 (2002).
- Royall, C. P., Louis, A. A. & Tanaka, H. Measuring colloidal interactions with confocal microscopy. *J. Chem. Phys.* **127**, 044507 (2007).
- Miller, M. A. & Frenkel, D. Competition of percolation and phase separation in a fluid of adhesive hard spheres. *Phys. Rev. Lett.* **90**, 135702 (2003).
- Sciortino, F. & Tartaglia, P. Structure factor scaling during irreversible cluster-cluster aggregation. *Phys. Rev. Lett.* **74**, 282–285 (1995).
- Furukawa, H. A dynamic scaling assumption for phase separation. *Adv. Phys.* **34**, 703–750 (1985).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** P.J.L. thanks D. Maas, M. Christiansen and S. Raghavachary for assistance in producing the renderings and movies. This work was supported by NASA, the NSF, the Harvard MRSEC, MIUR-Prin and the Marie Curie Research and Training Network on Dynamical Arrested States of Soft Matter and Colloids.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.J.L. ([plu@fas.harvard.edu](mailto:plu@fas.harvard.edu)).



## METHODS

**Colloid sample preparation.** Following our previously reported procedure<sup>22,24</sup>, we equilibrate sterically stabilized colloidal spheres of polymethylmethacrylate (PMMA) with DiI<sub>C18</sub> fluorescent dye in a 5:1 (by mass) solvent mixture of bromocyclohexane (CXB, Aldrich) and decahydronaphthalene (DHN, Aldrich) for several months. We add tetrabutylammonium chloride (TBAC, Fluke) until saturated (~4 mM) to screen long-range Coulombic repulsion. We then split the colloid suspension to create two stock solutions, adding linear polystyrene (Polymer Labs) depletant polymer to one. We buoyancy-match each stock solution individually by adding either CXB or DHN dropwise until particles remain neutrally buoyant after centrifuging at 1,000g for 30 min at 25.0 ± 0.1 °C. Mixing various ratios of the two stock solutions generates samples at varying  $c_p$ , while maintaining constant  $\phi$ , TBAC concentration, and buoyancy match.

We determine the radius  $a = 560 \pm 10$  nm of our particles with dynamic light scattering<sup>31</sup>. The solvent has viscosity  $\eta = 1.96$  mPa s at 25.0 ± 0.1 °C, measured with a Cannon-Fenske viscometer. For the depletant polystyrene, we selected two molecular weights,  $M_w = 69.2$  kDa and  $M_w = 681$  kDa. From Zimm plots of static light scattering data, we determine the radii of gyration  $r_g$  of the two polymers to be 10.0 and 33.0 nm, respectively. This yields  $\xi \equiv r_g/a$  of 0.018 and 0.059, respectively, and overlap concentrations  $c_p^* \equiv 3M_w/4\pi r_g^3 N_A$  of 27.2 and 7.5 mg ml<sup>-1</sup>, respectively, where  $N_A$  is Avogadro's number. In all cases, we directly measure the raw polymer concentrations as a mass ratio of mg polystyrene per g of total sample mass, which we express as a  $\phi$ -dependent free-volume  $c_p$  (mg ml<sup>-1</sup>) according to ref. 32.

**Confocal microscopy.** Following our previously reported imaging protocol<sup>22,24</sup>, we load each sample into a glass capillary of internal dimension 50 × 2 × 0.1 mm<sup>3</sup> (VitroCom), along with a small piece of magnetic wire with 25 μm diameter; we then seal the capillary with 5-min epoxy (DevCon). After sealing, we can rehomogenize the sample at any time by agitating the magnetic wire with a magnetic stirrer. We maintain the temperature of the microscope stage and surrounding air at 25.0 ± 0.2 °C, yielding a buoyancy match between colloid and solvent that is better than 10<sup>-4</sup>. With the confocal microscope, we collect 3D stacks of 181 8-bit images, each 1,000 × 1,000 pixels, at 10 frames per second. Each image stack covers a volume of 60 × 60 × 60 μm<sup>3</sup>, taken from the centre of the sample at least 20 μm away from any capillary surface to minimize edge effects.

Although larger clusters persist in these samples, the confocal microscope can collect 3D stacks only a few times a minute, far too slowly to track monomers, dimers and other small clusters. Therefore, to ensure a broad sampling, after homogenization and equilibration for four hours, we collect 26 independent 3D image stacks within each fluid sample, separated by 100 μm laterally, using our automated confocal microscope<sup>24</sup>. To observe the evolution of gel samples, we homogenize and immediately start observations, collecting 3D stacks of the same sample volume every 50 s for the first 5,000 s, then every 1,000 s for the next 100,000 s. In each 3D stack, we determine the 3D position of each particle more than 1 μm from the boundary of the imaging volume using previously described image-processing software<sup>24</sup>, and measure  $\phi$  for each sample from these particle counts. In total, we collected half a terabyte of image data and determined the positions of ~10<sup>8</sup> particles. Our 3D reconstructions were rendered with Pixar's RenderMan.

**Simulations.** We perform simulations of  $N = 10,000$  particles in a cubic box with periodic boundary conditions. For comparison to experimental samples with  $c_p = 0$ , we use the hard-sphere potential. For comparison to fluid samples with  $c_p \geq 0$ , we use three different attractive potential shapes, as shown in Fig. 2b: a square-well of width 0.04 $a$ , an Asakura–Oosawa potential<sup>33</sup> of maximum width 0.08 $a$ , and a generalized  $2\alpha$ - $\alpha$  Lennard–Jones potential with exponent  $\alpha = 100$  (ref. 34). For the Asakura–Oosawa potential, we use Monte Carlo simulations<sup>35</sup>; for the hard-sphere and square-well potentials, a standard event-driven algorithm<sup>36</sup>; and for the Lennard–Jones potential, molecular dynamics<sup>35</sup>. In the latter cases, the system is at first equilibrated in the NVT ensemble, followed by a

production run in the NVE ensemble, where 100 independent realizations are collected and analysed.

**Cluster mass distribution comparisons.** In particle configurations from both experiment and simulation, we define two particles as bonded if their centres are separated less than the bond distance  $r_b$ . All particles in a cluster share at least one bond with at least one other particle in the same cluster. Particles in one cluster share no bonds with particles in other clusters. Experimental uncertainties in particle locations arise from particle diffusion during confocal imaging, forcing the choice of  $r_b$  to be slightly larger than its ideal value of the particle diameter  $d = 2a$  plus the interaction range, for example, 1.08 $d$  for the previously described Asakura–Oosawa potential. We therefore set  $r_b$  by matching the hard-sphere simulations to the sample with  $c_p = 0$ , fixing this value for all samples at  $r_b = 1.16d$ ;  $n(s)$  comparisons are independent of the particular choice of  $r_b$ , so long as a consistent definition is applied to both experiments and simulations. For each experimental sample, we ran the simulations at the same  $\phi$ . The least-squares procedure to match  $n(s)$  from experiment and simulation equally weights all clusters.

**Static structure factor.** For fluid samples, we average the static structure factor  $S(q) \equiv \left\langle \left| \sum_{j=1}^N \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right|^2 \right\rangle / N$ , where  $\mathbf{r}_j$  are the coordinates of particle  $j$ , over the 26 independent configurations. For the gel samples, we follow a single configuration over time. We calculate  $S(q)$  for all particles more than 4 μm away from all boundaries of the imaging volume to minimize edge effects, which, if present, would affect only the range  $2qa \leq 0.2$ . For the first moment  $q_1(t) \equiv (\int_0^q S(q,t) q dq) / (\int_0^q S(q,t) dq)$ , we select the cut-off value  $2q_c a = 3$  to ensure the inclusion of all large wavelength contributions.

**Estimation of  $\phi$  and  $B_2^*$  for gel samples.** We extend the linear fit of the  $U/k_B T$  versus  $c_p$  for the fluid samples into the gel region at each  $\phi$  to estimate  $\tau \equiv 1/4(B_2^* - 1)$  for the gel samples shown in Fig. 4d. We estimate  $\phi_g$ , the internal volume fraction for spanning gel clusters, defined as those touching opposite faces of the cubic imaging volume, by measuring the free volume accessible to a spherical test particle of radius  $a$ . Splitting the imaging volume into a fine grid of cubes with edge length  $l_c \ll a$ , we place a test particle in each cube, and if no part of it intersects with spanning cluster particles, the volume occupied by the test particle is considered to be in the free volume. The fraction of sample volume not part of the free volume is considered to be the total cluster volume. The total volume of the particles within the cluster is their number times the volume per particle; dividing this by the total cluster volume yields  $\phi_g$ . We selected  $l_c = 0.25a$ , but the measured  $\phi_g$  values do not depend on  $l_c$  for values below  $\sim a/2$  and converge as expected for tests on standard structures, such as a cluster of the f.c.c. lattice, where  $\phi \rightarrow 0.74$ . This approach is strictly applicable only to structures, such as the present gels, where the solid phase is more dense at the scale of a single particle; our centrosymmetric interparticle attraction allows bond rotation without energy cost, thereby requiring multiple bonds for stable structures, leading to locally higher densities at the single-particle scale. By contrast, in the  $\phi \rightarrow 0$  limit of DLCA, the permanent particle bonds are fixed and do not allow rotation, resulting in a more string-like local structure. For a straight line of spheres, our measure yields the analytic result  $\phi = 4/(10 - \pi\sqrt{3}) \approx 0.88$ , but is less meaningful in this regime.

- Frissen, B. J. Revisiting the method of cumulants for the analysis of dynamic light-scattering data. *Appl. Opt.* **40**, 4087–4091 (2001).
- Lekkerkerker, H. N. W., Poon, W. C. K., Pusey, P. N., Stroobants, A. & Warren, P. B. Phase behavior of colloid + polymer mixtures. *Europhys. Lett.* **20**, 559–564 (1992).
- Asakura, S. & Oosawa, F. On interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.* **22**, 1255–1256 (1954).
- Vliegenthart, G. A., Lodge, J. F. M. & Lekkerkerker, H. N. W. Strong weak and metastable liquids structural and dynamical aspects of the liquid state. *Physica A* **263**, 378–388 (1999).
- Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford Univ. Press, Oxford, UK, 1989).
- Rapaport, D. C. *The Art of Molecular Dynamic Simulation* (Cambridge Univ. Press, Cambridge, UK, 1995).

## LETTERS

# Triple oxygen isotope evidence for elevated CO<sub>2</sub> levels after a Neoproterozoic glaciation

Huiming Bao<sup>1</sup>, J. R. Lyons<sup>2</sup> & Chuanming Zhou<sup>3</sup>

Understanding the composition of the atmosphere over geological time is critical to understanding the history of the Earth system, as the atmosphere is closely linked to the lithosphere, hydrosphere and biosphere. Although much of the history of the lithosphere and hydrosphere is contained in rock and mineral records, corresponding information about the atmosphere is scarce and elusive owing to the lack of direct records. Geologists have used sedimentary minerals, fossils and geochemical models to place constraints on the concentrations of carbon dioxide, oxygen or methane in the past<sup>1–4</sup>. Here we show that the triple oxygen isotope composition of sulphate from ancient evaporites and barites shows variable negative oxygen-17 isotope anomalies over the past 750 million years. We propose that these anomalies track those of atmospheric oxygen and in turn reflect the partial pressure of carbon dioxide ( $p_{\text{CO}_2}$ ) in the past through a photochemical reaction network linking stratospheric ozone to carbon dioxide and to oxygen<sup>5,6</sup>. Our results suggest that  $p_{\text{CO}_2}$  was much higher in the early Cambrian than in younger eras, agreeing with previous modelling results<sup>2</sup>. We also find that the  $^{17}\text{O}$  isotope anomalies of barites from Marinoan (~635 million years ago) cap carbonates display a distinct negative spike (around  $-0.70\text{‰}$ ), suggesting that by the time barite was precipitating in the immediate aftermath of a Neoproterozoic global glaciation, the  $p_{\text{CO}_2}$  was at its highest level in the past 750 million years. Our finding is consistent with the ‘snowball Earth’ hypothesis<sup>7,8</sup> and/or a massive methane release<sup>9</sup> after the Marinoan glaciation.

Since the discovery of widespread sulphate  $^{17}\text{O}$  anomalies in Earth’s continental deposits<sup>10</sup>, a considerable amount of triple oxygen isotope data has been gathered for sulphate of diverse origins<sup>11–14</sup>. Without exception, the  $^{17}\text{O}$  anomalies have been positive, as measured by  $\Delta^{17}\text{O}$  ( $\equiv \delta^{17}\text{O} - 0.52 \times \delta^{18}\text{O}$ ), in which  $\delta' \equiv \ln(R_{\text{sample}}/R_{\text{standard}})$  and  $R$  is the ratio of  $^{18}\text{O}/^{16}\text{O}$  or  $^{17}\text{O}/^{16}\text{O}$  (Supplementary Information 1–5). A positive anomaly indicates enrichment in  $^{17}\text{O}$  content with respect to what is expected from a terrestrial fractionation line. It is known that positive sulphate  $\Delta^{17}\text{O}$  is transferred from atmospheric ozone ( $\text{O}_3$ ) during oxidation of sulphur compounds in the atmosphere<sup>15</sup>. Overall, the  $\Delta^{17}\text{O}$  for terrestrial sulphate reaches as high as  $+5.84\text{‰}$ <sup>11</sup>.

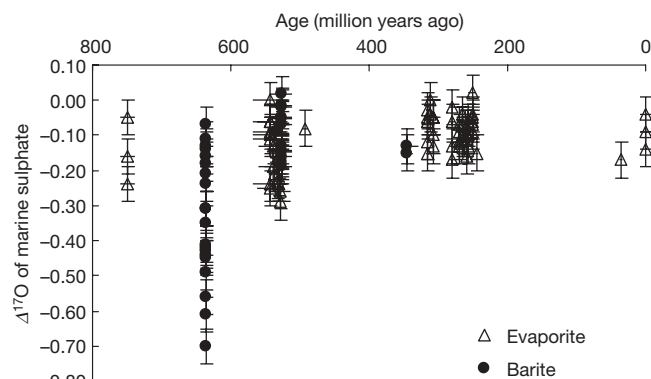
Here we report variable negative  $^{17}\text{O}$  anomalies among sulphate deposited on Earth surfaces over the past 750 million years (Fig. 1). Despite their small magnitudes, the negative  $^{17}\text{O}$  anomalies are larger than the analytical error of  $\pm 0.05\text{‰}$  and are reproducible. They were first noticed in an earlier survey of evaporite sulphate that had no direct link to atmospheric  $\text{O}_3$  chemistry<sup>16</sup>. An expanded survey revealed that marine evaporites and barites have  $\Delta^{17}\text{O}$  values ranging from  $+0.02$  to  $-0.70\text{‰}$ . Values more negative than  $-0.20\text{‰}$  are common in the Cambrian in diverse localities (for example, Siberia, Australia and India) whereas none occur in the late

Palaeozoic or modern settings. Most remarkably, barite cements from the Marinoan cap carbonate sequences deposited ~635 million years ago possess extremely negative  $\Delta^{17}\text{O}$  values both in West Africa (down to  $-0.40\text{‰}$ ) and in South China (down to  $-0.70\text{‰}$ ) (Supplementary Information 1).

We propose that the triple oxygen isotope compositions of sulphate carry a portion of the tropospheric  $\text{O}_2$  signal, which has had variable negative  $\Delta^{17}\text{O}$  values that are determined largely by stratospheric  $\text{O}_3$ – $\text{CO}_2$ – $\text{O}_2$  chemistry, and consequently the  $p_{\text{CO}_2}$  in the past (Fig. 2). We describe the connection between  $p_{\text{CO}_2}$  and sulphate  $\Delta^{17}\text{O}$  in three steps.

First, sulphate derived from oxidative weathering carries an atmospheric  $\text{O}_2$  signal. Sulphate oxygen ultimately comes from water and oxidants such as  $\text{O}_3$ ,  $\text{H}_2\text{O}_2$  or atmospheric  $\text{O}_2$ . When atmospheric  $\text{O}_3$  or  $\text{H}_2\text{O}_2$  is the oxidant, the product sulphate bears a positive  $\Delta^{17}\text{O}$  value<sup>10,15</sup>. The contribution of this atmospheric sulphate to ocean sulphate, however, is negligible compared with the influx from oxidative weathering. Early studies on sulphate derived from surface sulphur oxidations showed that, depending on reaction conditions and pathways, 0% to ~50% of the sulphate oxygen carries an  $\text{O}_2$  signal<sup>17</sup>. More recent laboratory experiments show that 8 to 15% of the oxygen in product sulphate came from  $\text{O}_2$  when  $\text{O}_2$  is involved in pyrite oxidation<sup>18</sup>.

Second, modern atmospheric  $\text{O}_2$  is known to have a small negative  $^{17}\text{O}$  anomaly, as first suggested by Bender *et al.*<sup>19</sup> and later documented by terrarium experiments<sup>20</sup>. The key chemical processes that bring a negative  $^{17}\text{O}$  anomaly to atmospheric  $\text{O}_2$  are the reactions of  $\text{O}_3$ – $\text{CO}_2$ – $\text{O}_2$  in the stratosphere<sup>5,6</sup>. The Chapman reactions produce  $\text{O}_3$  that is highly positive in both its  $\delta^{18}\text{O}$  and  $\Delta^{17}\text{O}$  values. Photolysis of  $\text{O}_3$  yields  $\text{O}(^1\text{D})$  which transfers the oxygen isotope



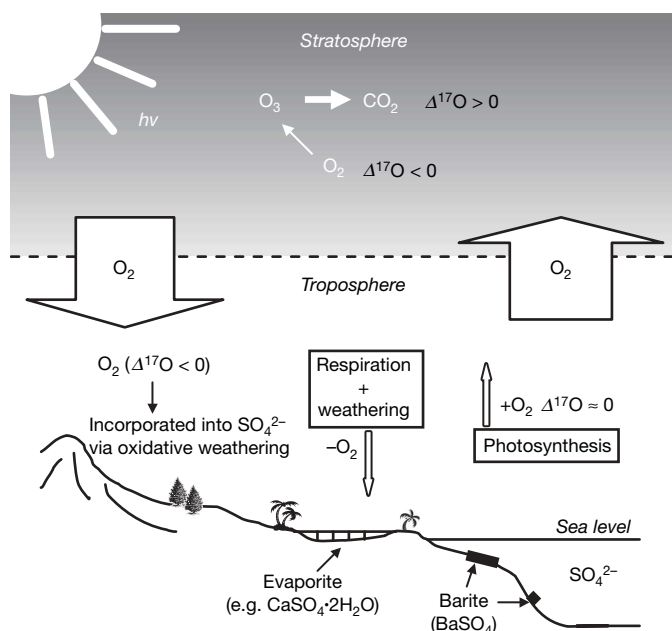
**Figure 1 | The  $\Delta^{17}\text{O}$  of evaporite and barite sulphate over the past 750 million years.**

<sup>1</sup>Department of Geology and Geophysics, E235 Howe-Russell Geoscience Complex, Louisiana State University, Baton Rouge, Louisiana 70803, USA. <sup>2</sup>Institute of Geophysics and Planetary Physics, Department of Earth and Space Sciences, University of California, Los Angeles, California 90095, USA. <sup>3</sup>State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, 39 East Beijing Road, Nanjing 210008, China.

signature of  $O_3$  to  $CO_2$  via the exchange reaction  $O(^1D) + CO_2 \rightarrow CO_3^* \rightarrow O(^3P) + CO_2$ . The net result is the generation of  $\delta^{18}O$ -high and  $\Delta^{17}O$ -positive  $CO_2$  and  $\delta^{18}O$ -low and  $\Delta^{17}O$ -negative  $O_2$ . The stratospheric  $O_2$  mixes with  $O_2$  in the troposphere but does not exchange oxygen with surface water. The exact magnitude of the non-mass-dependent  $^{17}O$  anomaly for current atmospheric  $O_2$  is debated, because of an uncertainty in kinetic fractionation slope<sup>21–23</sup>. Subtracting the effect of a newly calibrated slope<sup>23</sup> that is slightly different from the defined slope of 0.52, we calculated that tropospheric  $O_2$  has  $\sim 83\%$  of its  $\Delta^{17}O$  signal (around  $-0.19\%$ ) inherited from a stratospheric photochemical process (Supplementary Information 2 and 3.1). In fact,  $O_2$  is the only known atmospheric oxidant that has a negative  $^{17}O$  anomaly or a negative  $\Delta^{17}O$  ( $O_2$ ) value.

Third, the magnitude of  $\Delta^{17}O$  ( $O_2$ ) should have changed over time. This is because its controlling factors, such as  $p_{CO_2}$ ,  $p_{O_2}$ ,  $p_{O_3}$ , and the rates of photolysis, photosynthesis or respiration<sup>5,6,24</sup>, have not remained constant over geological time. For tropospheric  $O_2$ , the negative  $\Delta^{17}O$  signal is in a steady state, with influx from the stratosphere and photosynthesis, and outflux from respiration and oxidative weathering<sup>20</sup>. In the current atmosphere, the  $O_2$  flux from photosynthesis or respiration/weathering is only 1/300 of that from troposphere–stratosphere exchange<sup>20</sup>. Thus, the  $\Delta^{17}O$  of tropospheric  $O_2$  is determined by stratospheric  $O_2$  transferred into the troposphere during stratosphere–troposphere exchange, and the lifetime of tropospheric  $O_2$  with respect to photosynthesis and respiration. Higher  $p_{CO_2}$  implies a greater reservoir of  $\Delta^{17}O$ -positive  $CO_2$ , and a corresponding more negative  $\Delta^{17}O$  for  $O_2$ .

We ran a one-dimensional photochemical model of the modern atmosphere for a variety of  $p_{CO_2}$  conditions and for present-day  $p_{O_2}$  (see Methods Summary). The resulting  $\Delta^{17}O$  ( $O_2$ ) varies linearly with  $p_{CO_2}$  for small increases in  $p_{CO_2}$ , confirming the linear scaling assumption of Luz *et al.*<sup>20</sup>. The linear relationship is also confirmed by examining actual measurements of  $p_{CO_2}$  and  $\Delta^{17}O$  ( $O_2$ ) from two ice cores (GISP2 and Siple Dome) by two separate research groups<sup>20,24</sup>. According to the model the scaling becomes weaker than linear at high  $p_{CO_2}$  (Supplementary Information 6). An increase in the photosynthesis flux will reduce the magnitude of  $\Delta^{17}O$  ( $O_2$ ) by



**Figure 2 | How evaporite or barite sulphate records the negative  $^{17}O$  anomaly of tropospheric  $O_2$  that originated in the stratosphere.** The  $O_2$ – $O_3$ – $CO_2$  reaction network refers to those depicted in Supplementary Information 6, Supplementary Fig. 2. Sizes of the solid white arrows indicate relative fluxes (not to scale).

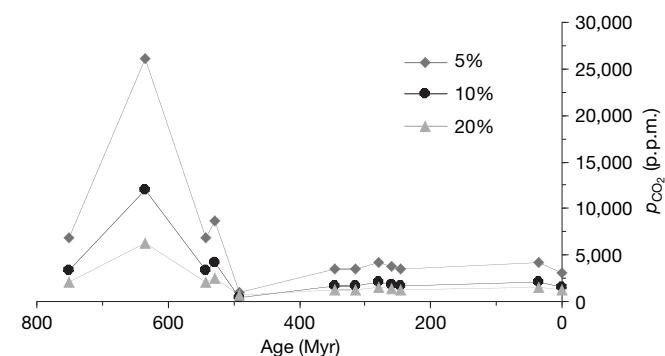
decreasing the atmospheric lifetime of  $O_2$  for a given  $O_2$  reservoir mass. To first order, a doubling of the photosynthesis flux will result in a halving of  $\Delta^{17}O$  ( $O_2$ ), assuming photosynthesis and respiration are in steady state.

A key attribute of sulphate is that it does not exchange oxygen with ambient water in most Earth surface environments<sup>25</sup>. Therefore, evaporite and barite sulphates from a given geological period would have a range of  $\Delta^{17}O$  values owing to initial variable sulphide oxidation pathways, subsequent redox cycling, or mixing of freshwater or atmospheric sulphate. However, the minimum  $\Delta^{17}O$  value for that period has to be set by the  $\Delta^{17}O$  ( $O_2$ ) at that time.

There are still temporal blanks to be filled in Fig. 1, but the low  $\Delta^{17}O$  cluster in the early Cambrian stands out as a distinct feature. The lowest sulphate  $\Delta^{17}O$  value,  $-0.29\%$ , is around three times more negative than those in the late Palaeozoic or modern evaporite and barite sulphates, suggesting a more negative  $\Delta^{17}O$  ( $O_2$ ) and thus a much higher atmospheric  $p_{CO_2}$  in Early Cambrian. This is consistent with the GEOCARBSULF modelling result on the  $p_{CO_2}$  and  $p_{O_2}$  history in the Phanerozoic<sup>2</sup>, in which  $p_{CO_2}$  was  $\sim 20$  times higher in the early Cambrian Period than in the late Palaeozoic or modern era. The most extraordinary feature in Fig. 1 is the barite cements from several  $\sim 635$ -Myr-old cap carbonate sequences in South China<sup>26</sup> and West Africa, occurring in dolostone matrices within a few metres above the Marinoan glacial diamictites. The  $\Delta^{17}O$  spike for these barites is around seven times more negative than that of the modern marine sulphate, implying that the atmosphere in the immediate aftermath of Marinoan deglaciation had the highest  $p_{CO_2}$  ratio ever for the past 750 million years.

Converting the record of marine sulphate  $\Delta^{17}O$  into a record of atmospheric  $p_{CO_2}$  requires us to know the fraction of sulphate oxygen contributed from atmospheric  $O_2$ . Comparing the average  $\Delta^{17}O$  for modern marine sulphate with  $\Delta^{17}O$  for modern  $O_2$  and considering the many uncertainties involved (Supplementary Information 5),  $\sim 10\% \pm 10\%$  of the marine sulphate oxygen  $\Delta^{17}O$  signal is estimated to have come from atmospheric  $O_2$ . Although consistent with a recent estimate<sup>18</sup>, this number is provisional and should be further calibrated both in the laboratory and in nature. Applying this value, we obtain a  $\Delta^{17}O$  ( $O_2$ ) value of  $-2.4\%$  and  $-6.5\%$  for the early Cambrian and at  $\sim 635$  Myr ago, respectively. According to our one-dimensional model, these  $\Delta^{17}O$  ( $O_2$ ) values imply  $p_{CO_2}$  of  $\sim 4,200$  p.p.m. in the early Cambrian and  $\sim 12,000$  p.p.m. at the time of barite precipitation in the Marinoan cap carbonate sequences. The estimated  $p_{CO_2}$  value is sensitive to the presumed  $O_2$  fraction in sulphate oxygen (Fig. 3).

It should be noted that the  $\sim 12,000$  p.p.m.  $p_{CO_2}$  may be a snapshot of a dynamic transition of atmospheric conditions immediately after the deglaciation<sup>27</sup>. The high  $p_{CO_2}$  could be the consequence of two causes, which are not mutually exclusive. First, the Neoproterozoic ‘snowball’ Earth hypothesis predicted a long-term build-up of



**Figure 3 | Model-calculated partial pressures of  $CO_2$  based on the lowest sulphate  $\Delta^{17}O$  value for a given period in geological history.** The diamonds, circles and triangles are for  $O_2$  signatures of 5, 10 and 20 mol% in sulphate oxygen, respectively.



volcanic CO<sub>2</sub> in the atmosphere, up to ~350 times the modern level, that finally offset the albedo effect and brought the Earth out of an otherwise perpetual snowball state<sup>7,8</sup>. Therefore, the high  $p_{\text{CO}_2}$  at the time of barite precipitation could be a relic of the high  $p_{\text{CO}_2}$  during the 'snowball' Earth. A second cause could be a catastrophic release of methane hydrates just after a global deglaciation, which, on the oxidation of methane, gave rise to a high  $p_{\text{CO}_2}$  level<sup>9</sup>. Whatever the cause, consistently more negative  $\Delta^{17}\text{O}$  values for the lower barite bed than for the upper one at multiple sites at Baizhu, Hubei Province, South China (Supplementary Information 7 and Supplementary Fig. 5) probably attest to a rapid CO<sub>2</sub> drawdown. The level of  $p_{\text{CO}_2}$  and the rate of CO<sub>2</sub> drawdown in the immediate aftermath of the Marinoan glaciation should ultimately be constructed with improved knowledge of seawater mixing and stratification in the post-glacial oceans and the exact fraction of O<sub>2</sub> being incorporated in marine sulphate.

Sulphate's negative  $\Delta^{17}\text{O}$  value is so far the only known mineral-isotope proxy that directly records the  $^{17}\text{O}$  anomaly of past atmospheric O<sub>2</sub>. Although the sulphate  $\Delta^{17}\text{O}$  record does not have the sensitivity to detect atmospheric  $p_{\text{CO}_2}$  changes between glacial and interglacial times, it can be particularly useful in evaluating extreme levels of atmospheric CO<sub>2</sub> or O<sub>2</sub> that have occurred throughout the Earth's history.

## METHODS SUMMARY

Sulphate was extracted from evaporites using Millipore water (18 M $\Omega$ ) and 1 M HCl, and precipitated as barite (BaSO<sub>4</sub>). Pure BaSO<sub>4</sub> was obtained from sedimentary barite via a modified DTPA (diethylenetriaminepentaacetic acid)-dissolution and re-precipitation (DDARP) method<sup>28</sup>. We extracted O<sub>2</sub> by using a CO<sub>2</sub>-laser fluorination system<sup>29</sup> and measured for the  $\delta^{17}\text{O}$  and  $\delta^{18}\text{O}$  on a Finnigan MAT 253 in a dual-inlet mode. The average O<sub>2</sub> sample size is ~25 micromoles, and is ~25% to 35% of the total barite oxygen yield. The standard deviation associated with the  $\Delta^{17}\text{O}$  is  $\pm 0.03\text{‰}$  for multiple ( $N \approx 3$ ) runs of the same O<sub>2</sub> gas on the MAT 253, and  $\pm 0.05\text{‰}$  for replicates of the same BaSO<sub>4</sub> via laser-fluorination. The reported  $\delta^{18}\text{O}$  was  $+9.4\text{‰}$  (a kinetic effect) plus the raw value obtained from the laser-fluorination method. The error is  $\pm 0.7\text{‰}$  for pure and fine-grained barite samples. Some  $\delta^{18}\text{O}$  values were obtained from the temperature-conversion elemental analyser with errors within  $\pm 0.5\text{‰}$ . No effort is made to distinguish them because an accurate  $\delta^{18}\text{O}$  value is not important to this study and errors in the  $\delta^{18}\text{O}$  and the  $\delta^{17}\text{O}$  cancel out during the  $\Delta^{17}\text{O}$  calculation. The sample was run only when the  $\Delta^{17}\text{O}$  zero enrichment was checked to be less than  $\pm 0.05\text{‰}$ . We have found that N<sub>2</sub> contamination can result in an erroneous, negative, but small  $\Delta^{17}\text{O}$  value when N<sub>2</sub>/O<sub>2</sub> is  $>2\%$  in our MAT 253. Most of our samples have N<sub>2</sub>/O<sub>2</sub>  $<0.3\%$  and none of the reported data had N<sub>2</sub>/O<sub>2</sub>  $>1\%$ .

In our one-dimensional photochemical model, temperature-dependent isotopomer-specific rate coefficients for O<sub>3</sub> formation<sup>30</sup> are used. Solar ultraviolet flux, vertical eddy mixing, temperature profiles, and atmospheric composition other than CO<sub>2</sub> are assumed to be unchanged relative to the present-day atmosphere. An O<sub>2</sub> atmospheric residence time of 1,200 yr<sup>20</sup> is used for all cases.

Received 7 September 2007; accepted 31 March 2008.

1. Rye, R., Kuo, P. H. & Holland, H. D. Atmospheric carbon dioxide concentrations before 2.2 billion years ago. *Nature* **378**, 603–605 (1995).
2. Berner, R. A. GEOCARBSULF: A combined model for Phanerozoic atmospheric O<sub>2</sub> and CO<sub>2</sub>. *Geochim. Cosmochim. Acta* **70**, 5653–5664 (2006).
3. Ekart, D. D., Cerling, T. E., Montanez, I. P. & Tabor, N. J. A 400 million year carbon isotope record of pedogenic carbonate: implications for paleoatmospheric carbon dioxide. *Am. J. Sci.* **299**, 805–827 (1999).
4. Royer, D. L. et al. Paleobotanical evidence for near present-day levels of atmospheric CO<sub>2</sub> during part of the Tertiary. *Science* **292**, 2310–2313 (2001).
5. Yung, Y. L., DeMore, W. B. & Pinto, J. P. Isotopic exchange between carbon dioxide and ozone via O(<sup>1</sup>D) in the stratosphere. *Geophys. Res. Lett.* **18**, 13–16 (1991).
6. Yung, Y. L. et al. Carbon dioxide in the atmosphere: Isotopic exchange with ozone and its use as a tracer in the middle atmosphere. *J. Geophys. Res.* **102** (D9), 10857–10866 (1997).
7. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball Earth. *Science* **281**, 1342–1346 (1998).
8. Pierrehumbert, R. T. High levels of atmospheric carbon dioxide necessary for the termination of global glaciation. *Nature* **429**, 646–649 (2004).
9. Jiang, G. Q., Kennedy, M. J. & Christie-Blick, N. Stable isotopic evidence for methane seeps in Neoproterozoic postglacial cap carbonates. *Nature* **426**, 822–826 (2003).

10. Bao, H. M. et al. Anomalous  $^{17}\text{O}$  compositions in massive sulphate deposits on the Earth. *Nature* **406**, 176–178 (2000).
11. Bao, H. M., Thieme, M. H., Loope, D. B. & Yuan, X. L. Sulfate oxygen-17 anomaly in an Oligocene ash bed in mid-North America: Was it the dry fogs? *Geophys. Res. Lett.* **30** (16), 1843, doi:10.1029/2003GL018689 (2003).
12. Howell, K. J. & Bao, H. M. Caliche as a geological repository for atmospheric sulfate. *Geophys. Res. Lett.* **33**, L13816, doi:10.1029/2006GL026518 (2006).
13. Alexander, B. et al. Climate driven changes in the oxidation pathways of atmospheric sulfur. *Geophys. Res. Lett.* **29** (14), 1685, doi:10.1029/2002GL014879 (2002).
14. Bao, H. M. & Marchant, D. R. Quantifying sulfate components and their variations in soils of the McMurdo Dry Valleys, Antarctica. *J. Geophys. Res.* **111** (D16301), doi:10.1029/2005JD006669 (2006).
15. Savarino, J., Lee, C. C. W. & Thieme, M. H. Laboratory oxygen isotopic study of sulfur (IV) oxidation: Origin of the mass-independent oxygen isotopic anomaly in atmospheric sulfates and sulfate mineral deposits on Earth. *J. Geophys. Res.* **105** (D23), 29079–29088 (2000).
16. Bao, H. M. Sulfate in modern playa settings and in ash beds in hyperarid deserts: Implication on the origin of  $^{17}\text{O}$ -anomalous sulfate in an Oligocene ash bed. *Chem. Geol.* **214**, 127–134 (2005).
17. van Stempvoort, D. R. & Krouse, H. R. in *Environmental Geochemistry of Sulfide Oxidation* Vol. 550 (eds Alpers, C. N. & Blowes, D. W.) 446–480 (American Chemical Society, Washington DC, 1994).
18. Balci, N., Shanks, W. C., Mayer, B. & Mandernack, K. W. Oxygen and sulfur isotope systematics of sulfate produced by bacterial and abiotic oxidation of pyrite. *Geochim. Cosmochim. Acta* **71**, 3796–3811 (2007).
19. Bender, M., Sowers, T. & Labeyrie, L. The Dole effect and its variations during the last 130,000 years as measured in the Vostok ice core. *Glob. Biogeochem. Cycles* **8**, 363–376 (1994).
20. Luz, B. et al. Triple-isotope composition of atmospheric oxygen as a tracer of biosphere productivity. *Nature* **400**, 547–550 (1999).
21. Young, E. D., Galy, A. & Nagahara, H. Kinetic and equilibrium mass-dependent isotope fractionation laws in nature and their geochemical and cosmochemical significance. *Geochim. Cosmochim. Acta* **66**, 1095–1104 (2002).
22. Miller, M. F. Isotopic fractionation and the quantification of  $^{17}\text{O}$  anomalies in the oxygen three-isotope system: an appraisal and geochemical significance. *Geochim. Cosmochim. Acta* **66**, 1881–1889 (2002).
23. Luz, B. & Barkan, E. The isotopic ratios  $^{17}\text{O}/^{16}\text{O}$  and  $^{18}\text{O}/^{16}\text{O}$  in molecular oxygen and their significance in biogeochemistry. *Geochim. Cosmochim. Acta* **69**, 1099–1110 (2005).
24. Blunier, T., Barnett, B., Bender, M. L. & Hendricks, M. B. Biological oxygen productivity during the last 60,000 years from triple oxygen isotope measurements. *Glob. Biogeochem. Cycles* **16** (3), 1029, doi:10.1029/2001GB001460 (2002).
25. Gamsjäger, H. & Murmann, R. K. in *Advances in Inorganic and Bioinorganic Mechanisms* Vol. 2 (ed. Sykes, A. G.) 317–381 (Academic, London, 1983).
26. Jiang, G. Q. et al. Stratigraphy, sedimentary structures, and textures of the late Neoproterozoic Doushantuo cap carbonate in South China. *J. Sediment. Res.* **76**, 978–995 (2006).
27. Kasemann, S. A. et al. Boron and calcium isotope composition in Neoproterozoic carbonate rocks from Namibia: evidence for extreme environmental change. *Earth Planet. Sci. Lett.* **231**, 73–86 (2005).
28. Bao, H. M. Purifying synthetic barite for oxygen isotope measurement by dissolution and reprecipitation in a chelating solution. *Anal. Chem.* **78**, 304–309 (2006).
29. Bao, H. M. & Thieme, M. H. Generation of O<sub>2</sub> from BaSO<sub>4</sub> using a CO<sub>2</sub>-laser fluorination system for simultaneous analysis of  $\delta^{18}\text{O}$  and  $\delta^{17}\text{O}$ . *Anal. Chem.* **72**, 4029–4032 (2000).
30. Mauersberger, K., Krankowsky, D., Janssen, C. & Schinke, R. in *Advances in Atomic, Molecular, and Optical Physics* Vol. 50 1–54 (Elsevier Academic, San Diego, 2005).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank numerous colleagues who have contributed evaporite and barite samples over the years, including W. T. Holser, M. Tucker, C. Spötl, R. Denison, C. Laughrey, S. C. Morris, S. P. Das Gupta, K. Benison, B.-H. Fu, S. Xiao, G. Retallack, W.-L. Zang, J. Hanor, B. Ellwood, D. Henry, B. Dutrow and A. J. Kaufman. One important barite sample, collected by G. P. Halverson, P. F. Hoffman and A. C. Maloof in Mauritania, West Africa, was obtained from M. H. Thieme and D. P. Schrag. We thank Huifeng Bao for field assistance, M. Khachatryan for laboratory assistance, and NSF, LSU, NASA (Planetary Atmospheres), the NNSF of China, the Chinese Academy of Sciences and the Chinese Ministry of Science and Technology for financial support.

**Author Contributions** H.B. designed the research, developed analytical procedures and performed measurements. J.R.L. did one-dimensional photochemical modelling and C.M.Z. directed fieldwork in South China. H.B. and J.R.L. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.B. (bao@lsu.edu).

# Seismogenic lavas and explosive eruption forecasting

Y. Lavallée<sup>1</sup>, P. G. Meredith<sup>2</sup>, D. B. Dingwell<sup>1</sup>, K.-U. Hess<sup>1</sup>, J. Wassermann<sup>1</sup>, B. Cordonnier<sup>1</sup>, A. Gerik<sup>1,3</sup> & J. H. Kruhl<sup>3</sup>

**Volcanic dome-building episodes commonly exhibit acceleration in both effusive discharge rate and seismicity before explosive eruptions<sup>1</sup>. This should enable the application of material failure forecasting methods to eruption forecasting<sup>2,3</sup>. To date, such methods have been based exclusively on the seismicity of the country rock<sup>4</sup>. It is clear, however, that the rheology and deformation rate of the lava ultimately dictate eruption style<sup>5</sup>. The highly crystalline lavas involved in these eruptions are pseudoplastic fluids that exhibit a strong component of shear thinning as their deformation accelerates across the ductile to brittle transition<sup>6</sup>. Thus, understanding the nature of the ductile–brittle transition in dome lavas may well hold the key to an accurate description of dome growth and stability. Here we present the results of rheological experiments with continuous microseismic monitoring, which reveal that dome lavas are seismogenic and that the character of the seismicity changes markedly across the ductile–brittle transition until complete brittle failure occurs at high strain rates. We conclude that magma seismicity, combined with failure forecasting methods, could potentially be applied successfully to dome-building eruptions for volcanic forecasting.**

Arc volcanoes commonly exhibit cycles of dome growth and collapse, leading sometimes to catastrophic explosions. Increasingly, these volcanoes are routinely monitored by multi-parameter (geophysical and geochemical) systems that provide a basis in practice for hazard management and forecasting of upcoming eruptions<sup>1</sup>. Fortunately for the monitoring process, precursory signals of volcanic unrest are common and numerous; yet their origins remain to be deciphered and properly characterized in a mechanistic way. In particular, volcanic eruptions generate various types of seismic signals, including continuous tremor, and it is within the complexities of their waveforms that the description of the responsible internal processes (for example, fluid oscillation, melt migration and fracturing) is likely to be found<sup>7–11</sup>. Although many doubts remain as to the exact nature of volcano–seismic source mechanisms, it is nevertheless commonly accepted that brittle failure along the conduit margin can play a major role<sup>12</sup>. To date, volcanic eruption forecasting models, such as the material failure forecast method (FFM), assume that the seismicity originates from fracturing of the volcanic edifice (and not from the magma)<sup>3,4</sup>. Recent fieldwork on eroded, shallow conduits has uncovered abundant evidence of a more complex magma rheology. In particular, structural and textural evidence have revealed the common existence of seismogenic fault zones in which multiple cycles of rupture, slip and healing have occurred in the magmas owing to strain rate variations across the glass transition<sup>13,14</sup>. Numerical models have further elucidated this shearing-induced fragmentation along the conduit walls; nevertheless, accurate modelling clearly awaits better rheological and seismological constraints<sup>15,16</sup>.

Ultimately it is the competition between the strain rate and the relaxation timescale of a melt that dictates whether the eruption will proceed effusively or explosively<sup>5</sup>. Classically, a pure, single-phase

melt behaves as a newtonian fluid at low strain rate, but as the deformation speeds up to near the relaxation timescale of the melt structure, the melt becomes non-newtonian. Viscous heating and microcracking ensue<sup>17,18</sup>. In nature, dome lavas inevitably contain variable amount of crystals and bubbles, yet the rheological influence of these features remains obscure<sup>19</sup>. Recent experimental and theoretical studies have helped in defining a realistic view of their non-newtonian behaviour<sup>6,20</sup>. Nevertheless their complex mechanical state, involving components of fluid and solid behaviours, denies us a complete constitutive relationship to date. Essentially, three effects have been recognized as the strain rate (or stress) is increased<sup>6</sup>. (1) An instantaneous viscosity decrease, recoverable upon stress release, defines multiphase lavas as pseudoplastic fluids with a strong component of shear thinning. As the strain rate is further increased, the viscosity becomes strain dependent; a delayed decrease in viscosity is accompanied by (2) viscous heating and (3) audible cracking. This late cracking of lavas, as it embraces the brittle regime, may hold the key to forecasting lava dome eruptions.

The experimental generation of cracks has been studied extensively in the field of rock mechanics<sup>21–23</sup>. Acoustic emissions generated by microcrack growth are used to track the development of macroscopic failure, as their temporal, spatial and size distribution follow a power law akin to that applicable to earthquakes<sup>23</sup>. Acoustic emission events are high-frequency strain waves analogous to low-frequency seismic waves in nature<sup>22</sup>. Yet, acoustic emission has seldom been used to characterize deformation of lavas, even though it has been proposed to provide “a sensitive procedure for monitoring the nature of creep deformation”<sup>24</sup>. The viscoelastic deformation described in our previous work is comparable to creep deformation<sup>6</sup>. Here we use acoustic emission for the first time (to our knowledge) to characterize the acoustic character of the non-newtonian regime of dome lavas across the ductile–brittle transition—from onset at low strain rate to failure at high strain rate—and to evaluate the failure prediction capability of the FFM.

The experimental arrangement for this investigation couples two now well-established techniques (see Supplementary Information). First, a well-calibrated, high-load, high-temperature uniaxial press was used to study the effects of stress and strain rate on the apparent viscosity of lavas from Colima (Mexico) and Bezymianny (Russia) volcanoes. Second, a fast acoustic-emission monitoring system was close-coupled to the press, and used to record acoustic-emission output simultaneously and continuously during each deformation experiment. Experiments were performed under stresses of 1–40 MPa and temperatures of 940–980 °C, that is, under the pressure–temperature conditions of dome-building eruptions.

Viscosity profiles for multiphase lavas deforming under successively increasing increments of stress have been described recently<sup>6</sup>. Here we extend that work to include the associated acoustic-emission energy released by microcracking during deformation (Fig. 1). Multiphase melt deformation under low stress (8 MPa) is typically characterized by a strong elasticity and thus a viscosity that increases

<sup>1</sup>Department of Earth and Environmental Sciences, Ludwig-Maximilians University, 80333 Munich, Germany. <sup>2</sup>Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. <sup>3</sup>Faculty of Civil and Geodetic Engineering, Technische Universität München, 80333 Munich, Germany.

at a decreasing rate until it stabilizes at a high, constant value (Fig. 1a). At this low stress, no viscous heating is generated and the temperature remains constant (Fig. 1b). A moderate number of acoustic-emission hits is recorded during the viscosity increase, but with time the acoustic-emission rate decreases to close to zero as viscosity stabilizes. As the acoustic-emission events are generally of low amplitude, the cumulative acoustic-emission energy also remains low (Fig. 1c). At intermediate stress (16 MPa), the viscosity is often observed to remain relatively constant over the duration of the deformation, and viscous heating sometimes increases the temperature (Fig. 1a, b). Under this regime, the acoustic-emission energy rate also remains essentially constant (Fig. 1c) but with occasional higher energy signals. Finally, at high stress (24 MPa), the viscosity decreases markedly during deformation (Fig. 1a). This extreme regime is characterized by 1.6 °C of viscous heating and an accelerating output of acoustic-emission energy (Fig. 1b, c). Overall, the

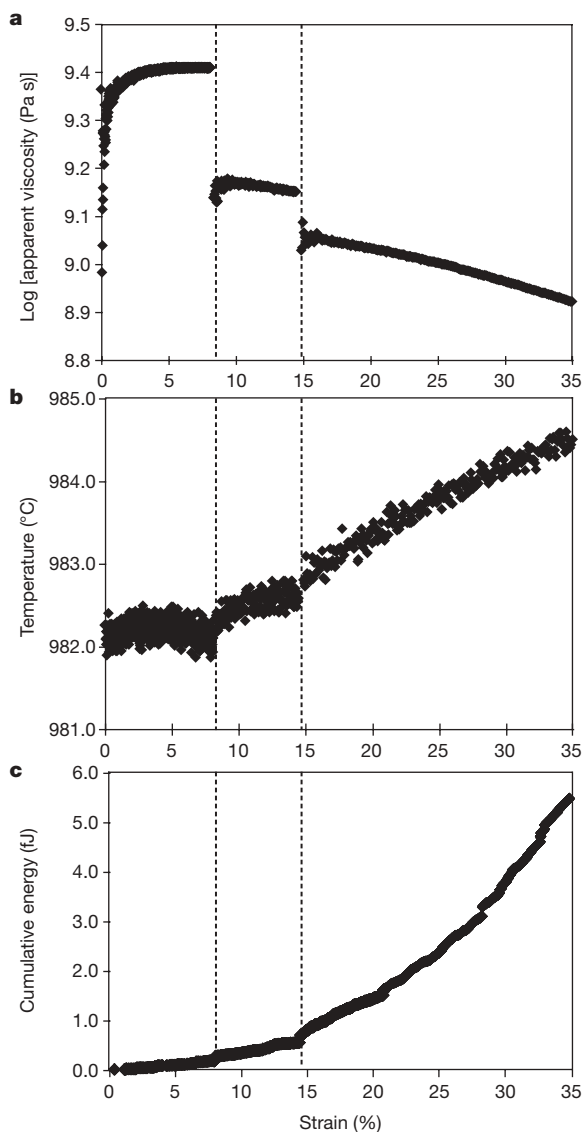
increase in acoustic-emission energy with increasing stress is due to an increase in both the number of hits and the individual hit amplitude (compare earthquake magnitude). This is, in turn, manifested in a decrease of the seismic *b*-value from >3.5 to as low as <1.5 in some cases (the *b*-value is the slope of the amplitude–frequency relationship; see Supplementary Information). This observation implies a change from more distributed small-scale cracking at lower stresses to more localized larger-scale cracking at higher stresses.

Suspension rheology involves cracking throughout the spectrum. The deformation is nearly aseismic at strain rates below  $10^{-4} \text{ s}^{-1}$ . Then, the rates of acoustic-emission output increases nonlinearly with increasing strain rate and accelerates as failure is approached (Fig. 2). The presence of crystals within a melt lowers the strain rate corresponding to the onset of the ductile–brittle transition in these multiphase magmas. Textural analysis of deformed samples indicates that cracking generally initiates in plagioclase crystals. At high strain rate, experiments revealed the alignment of crystals and the development of large-scale cracks (also reflected in the decrease of seismic *b*-value). Complementary quantitative analyses of fabrics developed in Colima and Bezmianny samples using the fabric analysis software AMOCADO<sup>25</sup> revealed an increase in the anisotropy—represented by the fitted ellipse's axial ratio—of the groundmass pattern by ~29% upon 33% strain (Fig. 3). The anisotropy of the crystal pattern, however, decreased by 19%. These observations suggest that during deformation, elongate crystals become broken into more equant fragments (lowering the crystal anisotropy) while the fragments from the original crystals align themselves perpendicularly to the applied stress to ease flow migration of the interstitial melt (increasing the overall anisotropy).

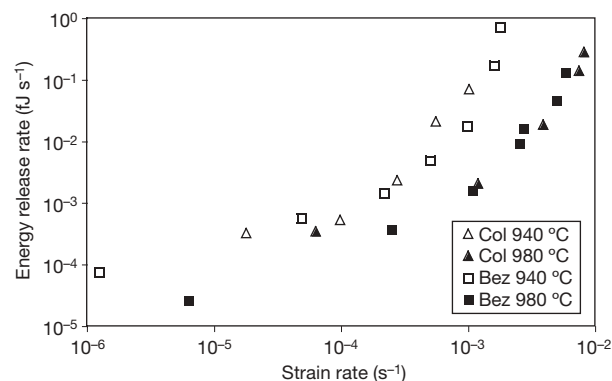
Given our observation that multiphase lavas behave in a brittle fashion at high strain rate, we have chosen to test whether crack growth and macroscopic failure of a multiphase melt at high strain rate is comparable to rock failure. The FFM relies on the production rate of precursory phenomena (for example, seismicity rate, acoustic-emission rate, seismic energy release), and correlating their accelerations to the likelihood of failure—in this case, of an eruption—via the equation

$$\frac{d^2\Omega}{dt^2} = A \left( \frac{d\Omega}{dt} \right)^\alpha \quad (1)$$

where  $d^2\Omega/dt^2$  and  $d\Omega/dt$  are the acceleration and rate of the phenomenon being monitored, and *A* and  $\alpha$  are empirically determined parameters<sup>2,3,26–28</sup>. More explicitly,  $\alpha$  is expected to evolve from 1 to 2 before an eruption<sup>27</sup>. A recent description of the fracturing time series that arise from random energy fluctuations within a finite volume subject to a constant remote stress proposed that the peaks in event

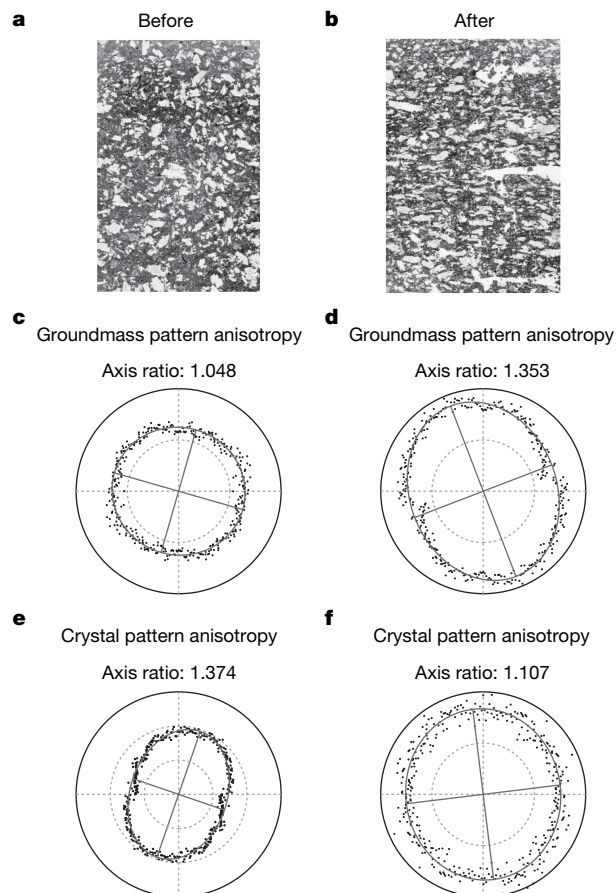


**Figure 1 | Experimental results for successive deformation of a Colima lava melt at 8, 16 and 24 MPa.** In each panel, the vertical dashed lines show when the pressure was changed; left line, 8 to 16 MPa; right line, 16 to 24 MPa. **a**, The apparent viscosity profile shows the instantaneous decrease associated with each stress increment. This is the origin of the non-newtonian behaviour. **b**, The internal melt temperature shows an increase associated with minor viscous heating at high stress. **c**, The cumulative acoustic-emission energy output is minor and constant at low to moderate stress, and increases exponentially at high stress (1 fJ =  $10^{-15}$  J).



**Figure 2 | Acoustic-emission energy release rates for Colima and Bezmianny lavas at different strain rates.** Although the crystallinities of Colima (Col; ~55% crystals) and Bezmianny (Bez; ~80% crystals) melt samples were significantly different, the behaviours of both melts were very similar at a given temperature. It is rather the temperature that may serve to attenuate acoustic emission.





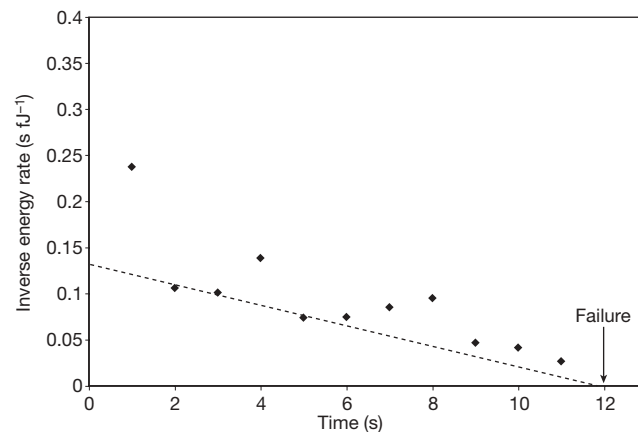
**Figure 3 | Anisotropy changes associated with deformation.** These images and anisotropy analyses show the results of an experiment on Colima lava at 940 °C under a pressure of 40 MPa. **a, b**, Photographs of thin sections before the experiment (**a**), and after the experiment (**b**; the applied stress was parallel to the long axis of the thin section). Both thin sections were prepared along the same plane in the original rock sample. Photograph **b** shows a clear alignment of the crystals perpendicular to the applied stress. **c, d**, The anisotropy of the groundmass increased by 39% when comparing the axis ratios of the fitted ellipse before the experiment (**c**), and after the experiment (**d**). **e, f**, In contrast, the anisotropy of the crystal pattern decreased by 19%, when comparing the results before (**e**) and after (**f**). These anisotropies are visualized by direction versus size coefficient distribution plots. The coefficients are calculated from segment length distributions received from scanlines superposed on the analysed fabric. Each ring is equivalent to a count of 500 units (**c**), 1,000 units (**d**), 1,250 units (**e**), and 50 units (**f**); see ref. 25 for details.

rate (rather than all seismic events) predict best the path to failure, and that  $\alpha = 2$  when approaching failure<sup>2</sup>. The equation can thus be simplified to:

$$\frac{1}{d\Omega/dt} = A(t_f - t) \quad (2)$$

where  $t_f$  is the expected time-to-failure. As the acceleration increases before failure, the extrapolation of the inverse rate to zero should provide the time-to-failure. Although empirically derived from the field of rock mechanics, this approach appears to provide a good representation of precursory accelerations preceding natural eruptions<sup>28</sup>—especially when the acceleration of energy released is used<sup>29</sup>. However, the predictions yielded by the model remain uncertain until shortly before an eruption, and thus an improved treatment must unfortunately await better rheological and seismological constraints<sup>29</sup>.

Our deformation experiments at very high strain rates on Colima lavas were characterized by an accelerating acoustic-emission event rate and energy rate until complete failure soon thereafter. We can



**Figure 4 | Application of the FFM to a Colima lava.** Experiment at 940 °C deformed under 40 MPa (strain rate of  $7 \times 10^{-3} \text{ s}^{-1}$ ). The FFM prediction was based on the extrapolation of peak energy rates (lower values on this inverse scale), following ref. 2. Extrapolation of peak energy rates after 4 s of deformation (dashed line) well predicts the time of complete failure (arrow).

therefore retrieve a data distribution analogous to acoustic emission measured for rocks before failure by simply inverting the acoustic-emission rate as shown in Fig. 4. Extrapolations of the peak energy rate data points after four seconds indeed yield a very accurate prediction of the macroscopic failure of lava, which occurred after 12 seconds. Although the test cannot be used to model more accurate  $\alpha$  values at this point, it strongly suggests that the choice of an exponent equalling two and the use of peak energy values are appropriate for the forecasts of lava dome eruption induced by shear strain. An earlier attempt to use the acceleration of seismic energy release to forecast two volcanic eruptions at Colima has shown that the method only became reasonably accurate shortly before the eruption<sup>28</sup>. That study further specifies that such forecasting models “require that the medium be considered as a closed-continuum system”. Under such conditions, our work raises the possibility of accurate early predictions. We attribute the difficulty of using the FFM in real time during volcanic crises to the use of seismic data that may not all originate from a common process. For instance, stick-slip motion along fault planes in the upper conduit (for example, Mount St Helens<sup>30</sup>) would alter the seismic signals derived from shear-induced fragmentation at greater depth. Such a signal distinction is an essential prerequisite for future forecasting attempts. The present findings indicate that run-away growth of the strain rate and seismic energy release rates before volcanic eruption is likely to be the result of lava crossing the ductile–brittle transition as a result of increasing strain rate.

The present work may have an effect on eruption forecast modelling. This series of rheological and acoustic tests has been able to expose the strongly seismogenic character of multiphase lavas across the ductile–brittle transitional field. At strain rates below  $10^{-4} \text{ s}^{-1}$ , lavas are nearly aseismic. In contrast, high-strain-rate experiments clearly reveal an accelerating rate of seismic energy release and a localization of the cracking until complete failure around  $10^{-3} \text{ s}^{-1}$  (at 940 °C) and  $10^{-2} \text{ s}^{-1}$  (at 980 °C). Energy rate acceleration before failure at high strain rates directly supports the application of FFM to dome-building eruptions. We conclude that it may be beneficial to test this technique in areas of volcanic unrest.

Received 15 November 2007; accepted 2 April 2008.

1. Sparks, R. S. J. Forecasting volcanic eruptions. *Earth Planet. Sci. Lett.* **210**, 1–15 (2003).
2. Kilburn, C. R. J. Multiscale fracturing as a key to forecasting volcanic eruptions. *J. Volcanol. Geotherm. Res.* **125**, 271–289 (2003).
3. Voight, B. A method for prediction of volcanic eruptions. *Nature* **332**, 125–130 (1988).
4. Voight, B. A relation to describe rate-dependent material failure. *Science* **243**, 200–203 (1989).
5. Dingwell, D. B. Volcanic dilemma: Flow or blow? *Science* **273**, 1054–1055 (1996).

6. Lavalée, Y., Hess, K. U., Cordonnier, B. & Dingwell, D. B. A non-Newtonian rheological law for highly-crystalline dome lavas. *Geology* **35**, 843–846 (2007).
7. McNutt, S. R. Volcanic seismology. *Annu. Rev. Earth Planet. Sci.* **33**, 461–491 (2005).
8. Neuberg, J., Luckett, R., Baptie, B. & Olsen, K. Models of tremor and low-frequency earthquake swarms on Montserrat. *J. Volcanol. Geotherm. Res.* **101**, 83–104 (2000).
9. Neuberg, J. Characteristics and causes of shallow seismicity in andesite volcanoes. *Phil. Trans. R. Soc. Lond. A* **358**, 1533–1546 (2000).
10. Chouet, B. A. Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
11. Harrington, R. M. & Brodsky, E. E. Volcanic hybrid earthquakes that are brittle-failure events. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL028714 (2007).
12. Neuberg, J. W. et al. The trigger mechanism of low-frequency earthquakes on Montserrat. *J. Volcanol. Geotherm. Res.* **153**, 37–50 (2006).
13. Tuffen, H. & Dingwell, D. B. Fault textures in volcanic conduits: Evidence for seismic trigger mechanisms during silicic eruptions. *Bull. Volcanol.* **67**, 370–387 (2005).
14. Tuffen, H., Dingwell, D. B. & Pinkerton, H. Repeated fracture and healing of silicic magma generate flow banding and earthquakes? *Geology* **31**, 1089–1092 (2003).
15. Gonnermann, H. M. & Manga, M. Explosive volcanism may not be an inevitable consequence of magma fragmentation. *Nature* **426**, 432–435 (2003).
16. Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1999).
17. Hess, K. U., Cordonnier, B., Lavalée, Y. & Dingwell, D. B. Viscous heating in rhyolite: an in situ determination. *Earth Planet. Sci. Lett.* (in the press).
18. Webb, S. L. & Dingwell, D. B. The onset of non-Newtonian rheology of silicate melts — a fiber elongation study. *Phys. Chem. Miner.* **17**, 125–132 (1990).
19. Petford, N. Rheology of granitic magmas during ascent and emplacement. *Annu. Rev. Earth Planet. Sci.* **31**, 399–427 (2003).
20. Costa, A. Viscosity of high crystal content melts: Dependence on solid fraction. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024303 (2005).
21. Prikryl, R., Lokajicek, T., Li, C. & Rudajev, V. Acoustic emission characteristics and failure of uniaxially stressed granitic rocks: The effect of rock fabric. *Rock Mech. Rock Eng.* **36**, 255–270 (2003).
22. Dobson, D. P., Meredith, P. G. & Boon, S. A. Detection and analysis of microseismicity in multi anvil experiments. *Phys. Earth Planet. Inter.* **143–44**, 337–346 (2004).
23. Ojala, I. O., Main, I. G. & Ngwenya, B. T. Strain rate and temperature dependence of Omori law scaling constants of AE data: Implications for earthquake foreshock-aftershock sequences. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020781 (2004).
24. Chmelik, F. et al. An evaluation of the creep characteristics of an AZ91 magnesium alloy composite using acoustic emission. *Mater. Sci. Eng. A* **338**, 1–7 (2002).
25. Gerik, A. & Kruhl, J. H. Towards automated pattern quantification: Time-efficient assessment of anisotropy of 2D patterns with AMOCADO. *Comput. Geosci.* (in the press).
26. Tokarev, P. On a possibility of forecasting of Bezymianny volcano eruptions according to seismic data. *Bull. Volcanol.* **26**, 379–386 (1963).
27. Cornelius, R. R. & Voight, B. Graphical and PC-software analysis of volcano eruption precursors according to the materials failure forecast method (FFM). *J. Volcanol. Geotherm. Res.* **64**, 295–320 (1995).
28. De la Cruz-Reyna, S. & Reyes-Davila, G. A. A model to describe precursory material-failure phenomena: Applications to short-term forecasting at Colima volcano, Mexico. *Bull. Volcanol.* **63**, 297–308 (2001).
29. Smith, R., Kilburn, C. R. J. & Sammonds, P. R. Rock fracture as a precursor to lava dome eruptions at Mount St Helens from June 1980 to October 1986. *Bull. Volcanol.* **69**, 681–693 (2007).
30. Iverson, R. M. et al. Dynamics of seismogenic volcanic extrusion at Mount St. Helens in 2004–05. *Geology* **444**, 439–443 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank O. Spieler for collecting the samples, K. T. Fehr, S. Bernstein and J. Pawlowski for assistance during microprobe analyses and M. Sieber for technical assistance. This is publication no. GEOTECH-315 of the research and development programme GEOTECHNOLOGIEN.

**Author Contributions** Y.L., P.G.M. and B.C. performed the acoustic-emission experiments; Y.L. analysed the data under the complementary supervisions of D.B.D., K.-U.H. and J.W.; and A.G. and J.H.K. performed the quantitative pattern analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.L. ([lavallee@min.uni-muenchen.de](mailto:lavallee@min.uni-muenchen.de)).

# Evidence for seismogenic fracture of silicic magma

Hugh Tuffen<sup>1,2</sup>, Rosanna Smith<sup>2</sup> & Peter R. Sammonds<sup>2</sup>

It has long been assumed that seismogenic faulting is confined to cool, brittle rocks, with a temperature upper limit of  $\sim 600^\circ\text{C}$  (ref. 1). This thinking underpins our understanding of volcanic earthquakes, which are assumed to occur in cold rocks surrounding moving magma. However, the recent discovery of abundant brittle–ductile fault textures in silicic lavas<sup>2–4</sup> has led to the counter-intuitive hypothesis that seismic events may be triggered by fracture and faulting within the erupting magma itself. This hypothesis is supported by recent observations of growing lava domes, where microearthquake swarms have coincided with the emplacement of gouge-covered lava spines<sup>5,6</sup>, leading to models of seismogenic stick-slip along shallow shear zones in the magma<sup>7</sup>. But can fracturing or faulting in high-temperature, eruptible magma really generate measurable seismic events? Here we deform high-temperature silica-rich magmas under simulated volcanic conditions in order to test the hypothesis that high-temperature magma fracture is seismogenic. The acoustic emissions recorded during experiments show that seismogenic rupture may occur in both crystal-rich and crystal-free silicic magmas at eruptive temperatures, extending the range of known conditions for seismogenic faulting.

Hundreds or thousands of small (magnitude  $M < 3$ ), low-frequency earthquakes occur during lava dome growth, typically tightly clustered around the conduit and dome  $< 2\text{ km}$  from the surface<sup>8–11</sup>. Events are commonly grouped in swarms, with similar waveforms indicating repeated activation of a near-static source<sup>4,9,10</sup>. The source mechanisms of these events have long been controversial, as strong attenuation in volcanic edifices makes full waveform inversions difficult and many potential mechanisms arise from the presence of interacting gas, liquid and solid phases<sup>12</sup>.

Researchers have recently recognized that small-scale brittle–ductile faults are abundant in silica-rich lavas<sup>3</sup> and display remarkably similar characteristics to tectonic faults inferred to have been seismogenic. This raises the possibility that syn-eruptive seismicity is triggered by a process analogous to tectonic faulting<sup>4</sup>. This trigger mechanism unifies existing, competing models, as faults nucleated by magma fracture<sup>2,13</sup> would involve stick- or creep-slip deformation<sup>7,14</sup>, while providing permeable pathways for transient escape of volcanic gases<sup>2,3</sup>.

The faulting hypothesis is further supported by recent observations of dome growth at Mount St Helens and Unzen, where shallow seismic swarms coincided with lava spine extrusion along

gouge-covered fault surfaces in the hot lava itself<sup>5,6</sup>. A growing number of researchers have thus proposed that fracturing of high-temperature, eruptible lava must control seismic triggering<sup>2–4,13,15</sup>, while also controlling the dynamics of dome emplacement<sup>7</sup> and degassing patterns<sup>16</sup>.

To test this hypothesis, we have done uniaxial and triaxial deformation experiments on samples of both glassy and crystalline lavas at temperatures up to  $900^\circ\text{C}$  (Table 1). The glassy lava was aphyric bubble-free rhyolitic obsidian from Krafla, Iceland (100% glass), and the crystalline lava was porphyritic andesite (21% phenocrysts  $< 2.5\text{ mm}$  long,  $< 2\%$  glass) from Mt Shasta, California. Further sample details are given in Supplementary Information.

Cylindrical samples  $75\text{ mm}$  in length and  $25\text{ mm}$  diameter, jacketed in a ductile iron sleeve, were deformed in compression in a high-pressure, high-temperature triaxial cell<sup>17</sup>. The sample dimensions greatly exceeded maximum crystal sizes, thus providing representative mechanical data. In triaxial tests, an all-round hydrostatic pressure was first applied to the sample and maintained at a set value (the ‘confining pressure’), and then the sample was heated and maintained at a set temperature using an internal heater. An axial load was applied to the rock sample by a  $200\text{-kN}$  servo-controlled pressure-balanced actuator at constant displacement rate (that is, constant strain rate). Acoustic emissions were detected continuously using a piezoelectric transducer attached to the loading piston via a waveguide. The use of a waveguide, which was essential to prevent high temperatures damaging the transducer, attenuates the acoustic signal but does not change the overall acoustic-emission frequency–magnitude relationships<sup>18</sup>. Samples were deformed at a range of constant strain rates (from  $10^{-4.3}$  to  $10^{-5}\text{ s}^{-1}$ , with total strains of  $\leq 4\%$ ) and temperatures in order to attain both brittle and ductile deformation behaviour (Table 1).

Figure 1 shows the results of deformation experiments done on obsidian at  $645^\circ\text{C}$ , close to its glass transition. At the higher strain rate of  $10^{-4.3}\text{ s}^{-1}$  (Fig. 1a), initial quasi-elastic loading was followed by brittle–ductile behaviour characterized by a sequence of small, abrupt stress drops and associated reduction in compliance, which indicates progressive damage in the sample<sup>19</sup>. There is a clear correlation between stress drops and bursts of acoustic emission shown by the steps in the cumulative acoustic energy release (Fig. 1a), which we attribute to cracking in the sample. The seismic  $b$ -value (the log-linear slope of the acoustic-emission frequency–magnitude

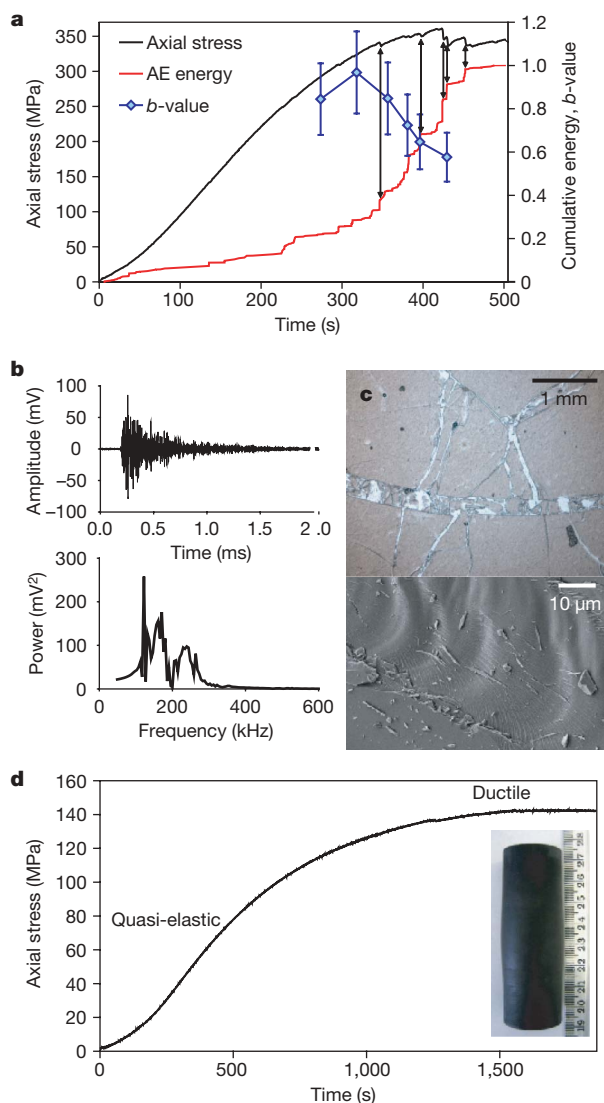
**Table 1 | Summary of experimental conditions**

| Sample | Material | Confining pressure (MPa) | Temperature ( $^\circ\text{C}$ ) | Strain rate ( $\text{s}^{-1}$ ) | Sample behaviour   |
|--------|----------|--------------------------|----------------------------------|---------------------------------|--|
| SA45   | Andesite | 0.3                      | 900                              | $10^{-5}$                       | Some ductile deformation, brittle shear failure            |
| SA43   | Andesite | 10                       | 900                              | $10^{-5}$                       | Predominantly ductile deformation with some shear cracking |
| SA42   | Andesite | 10                       | 600                              | $10^{-5}$                       | Elastic–brittle  |
| H15-3  | Obsidian | 0.3                      | 645                              | $10^{-4.3}$                     | Some ductile deformation, axial cracking                   |
| H6     | Obsidian | 0.3                      | 645                              | $10^{-4.9}$                     | Ductile barrelling   |
| H15-4  | Obsidian | 0.1                      | 20                               | $10^{-5}$                       | Elastic–brittle  |

<sup>1</sup>Department of Environmental Science, Lancaster University, Lancaster LA1 4YQ, UK. <sup>2</sup>Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK.



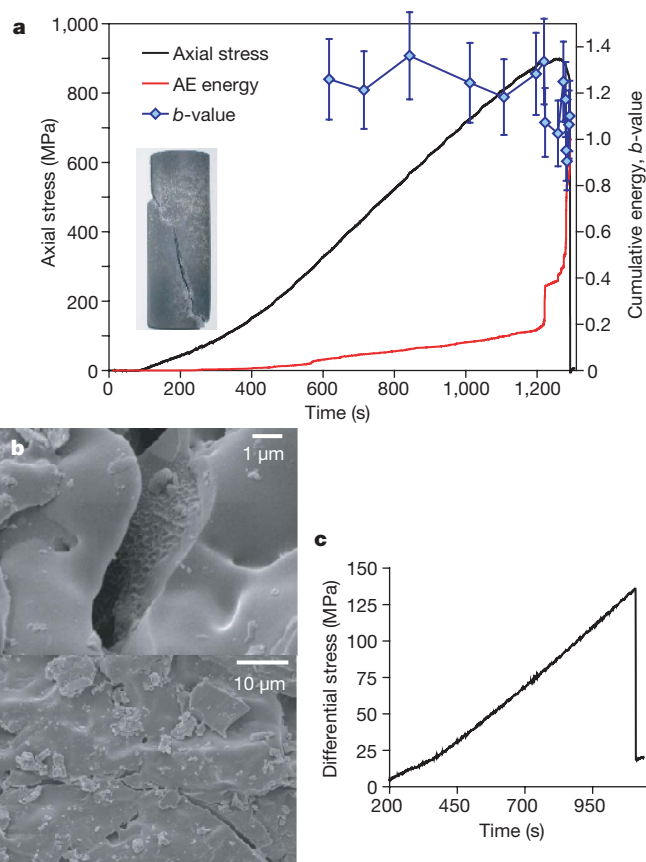
distribution, calculated here using Aki's maximum likelihood method<sup>20</sup>) decreases as the peak stress is approached. This decrease is indicative of microcrack extension and coalescence occurring with ongoing sample deformation<sup>21</sup>. A representative acoustic-emission waveform and power spectrum are shown in Fig. 1b. Energy is predominantly in the 100–300 kHz range. The onset is abrupt, and the waveform is typical of acoustic-emission events recorded during brittle failure of other crustal rock samples<sup>22</sup> and similar to waveforms we recorded during brittle failure of the obsidian at room temperature (Supplementary Information). Post-experiment sample analysis showed that numerous predominantly axial cracks had formed, with curved, conchoidal surfaces and local zones of cataclasis



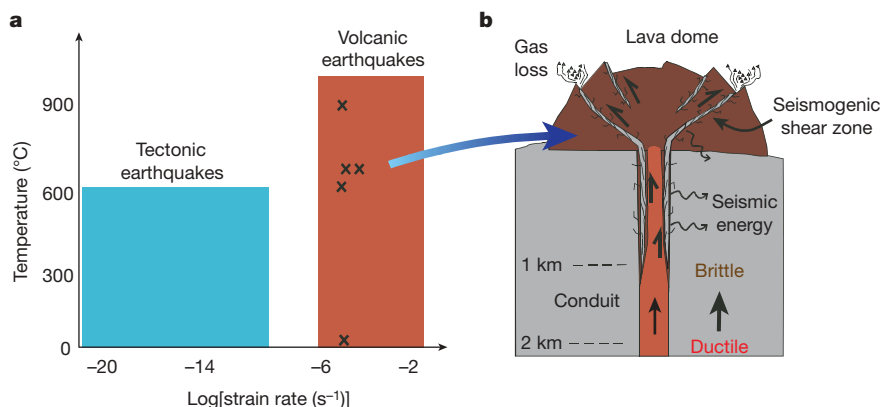
**Figure 1 | Experimental results from high-temperature fracture of rhyolitic obsidian.** **a**, Axial stress, normalized cumulative acoustic-emission (AE) energy, and acoustic-emission  $b$ -values against time for uniaxial deformation of rhyolitic obsidian at 645 °C and  $10^{-4.3} \text{ s}^{-1}$ . Jumps in cumulative energy correspond with stress drops (arrows) and drops in compliance, indicating that cracking of the sample is associated with release of acoustic energy. Error bars show 95% confidence limits. **b**, Waveform (top) and power spectrum (bottom) of a typical acoustic-emission event, showing the sharp onset and high frequency content (predominantly 100–300 kHz) that are characteristic of brittle failure. **c**, Photomicrograph (top) of post-experiment obsidian sample H15-3, sectioned normal to applied load, showing formation of gouge on curved brittle fracture surfaces; SEM image (bottom) showing detail of typical fracture surface. **d**, Loading behaviour of obsidian deformed at 645 °C and  $10^{-4.9} \text{ s}^{-1}$ , showing ductile deformation that led to barrelling of the sample H6 (inset; scale in cm).

where slip had occurred between major subparallel cracks (Fig. 1c, upper image). Scanning electron microscopy (SEM) images of surfaces (Fig. 1c, lower image) show micrometre-scale hackle markings typical of brittle glass fracture<sup>23</sup>. At lower strain rates of  $10^{-4.9} \text{ s}^{-1}$  (Fig. 1d), ductile deformation of the sample occurred after initial quasi-elastic loading, leading to sample barrelling but without crack formation. No acoustic emissions were detected during such ductile behaviour. The sustained peak stress of 140 MPa at this strain rate reflects the melt viscosity of  $\sim 10^{13} \text{ Pa s}$ . Such strongly strain-rate-dependent ductile–brittle behaviour is typical of silicate melts<sup>24</sup>.

Figure 2a shows the results of a deformation experiment on andesite at 900 °C. After a prolonged phase of quasi-elastic loading, the sample undergoes strain hardening ( $\sim 0.37\%$  strain) close to peak stress (of around 90 MPa) accompanied by strong acoustic-emission activity (Fig. 2a). Post-peak stress, the sample underwent a small but significant phase of 0.06% strain softening deformation and accelerating acoustic-emission activity, leading to dynamic failure. Failure involved the formation of a single shear fault at 17° to the loading axis. The  $b$ -value dropped to a local minimum at peak stress, recovered during strain softening deformation, before reaching a lower minimum as the sample failed. The  $b$ -value ‘double minimum’ is attributed to an increase in acoustic-emission amplitudes and crack tip stress intensities during pre-peak stress loading, which then



**Figure 2 | Experimental results from high-temperature fracture of Mt Shasta andesite.** **a**, Axial stress, normalized cumulative acoustic-emission energy, and acoustic-emission  $b$ -values against time for uniaxial deformation of Mt Shasta andesite (sample SA45) at 900 °C and  $10^{-5} \text{ s}^{-1}$ . The sudden increase in acoustic-emission energy occurs close to peak stress, and the  $b$ -value drops sharply at failure. The post-experiment sample displays a through-going shear fracture (inset). Error bars show 95% confidence limits. **b**, SEM images of fracture surfaces in andesite deformed at 900 °C. Top, brittle–ductile textures preserved in glass on the shear fracture plane in SA43; bottom, quenched melt on a fracture surface in sample SH45. **c**, Differential stress against time for triaxial deformation of Mt Shasta andesite (sample SA42) at 600 °C and  $10^{-5} \text{ s}^{-1}$  with 10 MPa confining pressure. Brittle failure occurs immediately after the peak stress.



**Figure 3 | Schematic diagrams comparing faulting in silicic magma to tectonic faulting and showing where fault zones may develop during a lava dome eruption. a**, Diagram showing the approximate range of temperatures and strain rates for seismogenic faulting in the lithosphere as a whole

dropped during strain softening deformation before increasing again at dynamic failure<sup>25</sup>. Post-experiment analysis using the SEM shows that brittle–ductile deformation of a melt phase had occurred on the fracture surface (Fig. 2b). In contrast, andesite samples deformed at 600 °C only exhibited 0.07% post-yield strain and failed immediately after reaching their peak stress (Fig. 2c).

Our experimental results extend the range of known conditions for seismogenic rupture to include magma at 900 °C (Fig. 3a). This is significantly greater than the 600 °C limit proposed for faulting elsewhere in the lithosphere<sup>1</sup>, owing to compositional effects (the high viscosity of silicic magmas) and abnormally high strain rates in magma (about ten orders of magnitude faster than the lithosphere as a whole).

Recent conduit flow models<sup>4,16</sup> have shown that the high viscosity of silicic magma makes it prone to shear fracture during ascent in the shallow conduit, owing to the rheological stiffening of silicic magma driven by shallow degassing and crystallization<sup>26</sup>. This is consistent with the widespread field evidence for localized strain on shear zones in silicic magma, either at the conduit walls<sup>2–4</sup> or bounding lava spines<sup>5,6</sup>. Similar structures do not develop in more basic magma, owing to its lower viscosity<sup>2</sup>.

Although volcano-tectonic earthquakes are generally thought to occur when cold rocks are fractured by moving magma<sup>9</sup>, acoustic-emission waveforms recorded during our high-temperature experiments on obsidian have sharp onsets similar to volcano-tectonic earthquakes and their frequency content can also be scaled to volcano-tectonic events. Given that dominant frequencies of earthquakes scale inversely with source dimension<sup>27</sup>,  $d \times f$  (source dimension times frequency) should be the same for volcanic earthquakes and acoustic emission recorded in the laboratory if they have the same mechanism<sup>28</sup>. For a typical  $M = 2$  volcano-tectonic earthquake,  $d = 100$  m and  $f = 15$  Hz, so  $d \times f = 1.5 \times 10^3$  m Hz; for the experiments,  $d = 0.01$  m and  $f = 150$  kHz, so  $d \times f = 1.5 \times 10^3$  m Hz, indicating the same brittle mechanism. The failure mode of experimental samples was predominantly tensile, whereas natural magmas often fail in shear<sup>2–4,13,16</sup>. However, the insight that the rupture is seismogenic is valid because shear failure typically generates more seismic energy than tensile failure for the same strain<sup>29</sup>.

Shallow seismicity during lava dome eruptions is, however, dominated by low-frequency events (hybrid and long-period earthquakes, with dominant energy in the 1–10 Hz range). Source inversions<sup>15</sup> have shown that these events may constitute brittle failure events<sup>2,13</sup>, and it has been speculated that the low frequencies reflect slow rupture of high-temperature magma<sup>3,30</sup>. Although we have not found any obvious difference in the frequency content of hot and cold failure events, further study is required to show whether rupture velocities show any temperature dependence. Alternatively, the low

(‘tectonic earthquakes’) and in silicic magma (‘volcanic earthquakes’). Crosses indicate the conditions of the experiments described in this paper. **b**, Cartoon indicating how seismogenic fault zones develop in magma at the conduit walls and within lava domes, and act as pathways for gas escape.

frequency content of hybrid and long-period earthquakes and their extended monochromatic codas may be attributed to conduit excitation<sup>4,8,9</sup> or path effects<sup>15</sup>. Magma fracture would, in this case, play a key role in creating transient permeable pathways for gas release and conduit excitation<sup>2–4</sup>.

As demonstrated by the striking similarity between fault textures in magma and cooler crustal rocks<sup>3,5</sup>, faulting in magma is analogous to seismogenic faulting elsewhere in the crust, despite occurring on dramatically shorter temporal and spatial scales (Fig. 3). Event magnitudes are limited by the dimensions of the magma body (hundreds of metres), and the lifetimes of individual faults are many orders of magnitude shorter than those of tectonic faults. Study of this hot, fast endmember of seismogenic faulting cycles may therefore shed light on faulting elsewhere in the lithosphere, as the evolution of short-lived swarms of similar events during dome eruptions<sup>10</sup> records the initiation and death of seismogenic fault systems.

Further experimentation is required to determine how seismic source characteristics and path effects (for example, rupture velocities and attenuation) relate to the mechanical state of the magma. This will greatly improve our understanding of how volcanic earthquakes relate to potentially hazardous changes in activity. It will also provide new insights into the mechanisms of a newly discovered type of seismogenic faulting.

Received 3 December 2007; accepted 2 April 2008.

- McKenzie, D., Jackson, J. & Priestley, K. Thermal structure of oceanic and continental lithosphere. *Earth Planet. Sci. Lett.* **233**, 337–349 (2005).
- Tuffen, H., Dingwell, D. B. & Pinkerton, H. Repeated fracture and healing of silicic magma generate flow banding and earthquakes? *Geology* **31**, 1089–1092 (2003).
- Tuffen, H. & Dingwell, D. B. Fault textures in volcanic conduits: Evidence for seismic trigger mechanisms during silicic eruptions. *Bull. Volcanol.* **67**, 370–387 (2005).
- Neuberg, J. *et al.* The trigger mechanism of low-frequency earthquakes on Mount St. Helens. *J. Volcanol. Geotherm. Res.* **153**, 37–50 (2006).
- Cashman, K. V., Thornber, C. R. & Pallister, J. S. In *A Volcano Rekindled: The First Year of Renewed Eruption at Mount St. Helens, 2004–2006* (eds Sherrod, D. R., Scott, W. E. & Stauffer, P. H.) Ch. 19 (USGS Professional Paper, US Geological Survey, Washington DC, in the press).
- Nakada, S., Shimizu, H. & Ohta, K. Overview of the 1990–1995 eruption at Unzen volcano. *J. Volcanol. Geotherm. Res.* **89**, 1–22 (1999).
- Iverson, R. M. *et al.* Dynamics of seismogenic volcanic extrusion at Mount St Helens in 2004–2005. *Nature* **444**, 439–443 (2006).
- Chouet, B. Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
- Neuberg, J. Characteristics and causes of seismicity in andesite volcanoes. *Phil. Trans. R. Soc. Lond.* **358**, 1533–1546 (2000).
- Umakoshi, K. *et al.* Seismic activity associated with the growth of the lava dome at Unzen volcano (November 1993 – January 1994) — grouping of earthquakes on the basis of cross-correlations among their waveforms. *Bull. Volcanol. Soc. Jpn* **47**, 43–55 (2002).

11. Smith, R., Kilburn, C. R. J. & Sammonds, P. R. Rock fracture as a precursor to lava dome eruptions at Mount St Helens from June 1980 to October 1986. *Bull. Volcanol.* **69**, 681–693 (2007).
12. McNutt, S. R. Volcanic seismology. *Annu. Rev. Earth Planet. Sci.* **33**, 461–491 (2005).
13. Goto, A. A new model for volcanic earthquake at Unzen volcano: Melt rupture model. *Geophys. Res. Lett.* **26**, 2541–2544 (1999).
14. Voight, B. *et al.* Magma flow instability and cyclic activity at Soufriere Hills volcano, Montserrat, British West Indies. *Science* **283**, 1138–1142 (1999).
15. Harrington, R. M. & Brodsky, E. E. Volcanic hybrid earthquakes that are brittle failure events. *Geophys. Res. Lett.* **34**, L06308, doi:10.1029/2006GL028714 (2007).
16. Gonnermann, H. M. & Manga, M. Explosive volcanism may not be an inevitable consequence of magma fragmentation. *Nature* **426**, 432–435 (2003).
17. Rocchi, V., Sammonds, P. R. & Kilburn, C. R. J. Fracturing of Etnean and Vesuvian rocks at high temperatures and low pressures. *J. Volcanol. Geotherm. Res.* **132**, 137–157 (2004).
18. Meredith, P. G. & Atkinson, B. K. Stress corrosion and acoustic emission during tensile crack propagation in Whin Sill dolerite and other basic rocks. *Geophys. J. R. Astron. Soc.* **75**, 1–21 (1983).
19. Main, I. G. A damage mechanics model for power-law creep and earthquake aftershock and foreshock sequences. *Geophys. J. Int.* **142**, 151–161 (2000).
20. Aki, K. Maximum likelihood estimate of  $b$  in the formula  $\log N = a - bM$  and its confidence limits. *Bull. Earthq. Res. Inst. Univ. Tokyo* **43**, 237–239 (1965).
21. Main, I. G., Sammonds, P. R. & Meredith, P. G. Application of a modified Griffith criterion to the evolution of fractal damage during compressional rock failure. *Geophys. J. Int.* **115**, 367–380 (1993).
22. Read, A. D., Ayling, M. R., Meredith, P. G. & Murrell, S. A. F. Microcracking during triaxial deformation of porous rocks monitored by changes in rock physical properties. II. Pore volumetry and acoustic emission measurements on water saturated rocks. *Tectonophysics* **245**, 223–235 (1995).
23. Hull, D. *Fractography* (Cambridge Univ. Press, Cambridge, UK, 1999).
24. Webb, S. L. & Dingwell, D. B. Non-Newtonian rheology of igneous melts at high stresses and strain rates: Experimental results for rhyolite, andesite, basalt and nephelinite. *J. Geophys. Res.* **95**, 15695–15701 (1990).
25. Sammonds, P. R., Meredith, P. G. & Main, I. G. Role of pore fluids in the generation of seismic precursors to shear fracture. *Nature* **359**, 228–230 (1992).
26. Sparks, R. S. J. *et al.* Control on the emplacement of the andesite lava dome of the Soufriere Hills volcano, Montserrat, by degassing-induced crystallization. *Terra Nova* **12**, 14–20 (2000).
27. Aki, K. & Richards, P. G. *Quantitative Seismology: Theory and Methods* (Freeman and Co., San Francisco, California, 1980).
28. Burlini, L. *et al.* Seismicity preceding volcanic eruptions: New experimental insights. *Geology* **35**, 183–186 (2007).
29. McGarr, A. Seismic moments and volume changes. *J. Geophys. Res.* **81**, 1487–1494 (1976).
30. Uchida, N. & Sakai, T. Analysis of peculiar volcanic earthquakes at Satsuma-Iojima volcano. *Earth Planets Space* **54**, 197–209 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to J. Bowles, S. Boon and N. Hughes for technical assistance with the experiments, C. Kilburn and H. Pinkerton for discussions and W. Hirt for help during fieldwork on Mt Shasta. This research was supported by the UK Natural Environment Research Council and the Leverhulme Trust (H.T.).

**Author Contributions** H.T. provided the initial idea, carried out deformation experiments on obsidian, analysed data and co-wrote the paper. R.S. carried out deformation experiments on andesite, analysed data and co-wrote the paper. P.R.S. co-wrote the paper and assisted with the experimental programme.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.T. ([h.tuffen@lancaster.ac.uk](mailto:h.tuffen@lancaster.ac.uk)).



# A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders

Jason S. Anderson<sup>1</sup>, Robert R. Reisz<sup>2</sup>, Diane Scott<sup>2</sup>, Nadia B. Fröbisch<sup>3</sup> & Stuart S. Sumida<sup>4</sup>

The origin of extant amphibians (Lissamphibia: frogs, salamanders and caecilians) is one of the most controversial questions in vertebrate evolution, owing to large morphological and temporal gaps in the fossil record<sup>1–3</sup>. Current discussions focus on three competing hypotheses: a monophyletic origin within either Temnospondyli<sup>4–7</sup> or Lepospondyli<sup>8–10</sup>, or a polyphyletic origin with frogs and salamanders arising among temnospondyls and caecilians among the lepospondyls<sup>11–16</sup>. Recent molecular analyses are also controversial, with estimations for the batrachian (frog–salamander) divergence significantly older than the palaeontological evidence supports<sup>17,18</sup>. Here we report the discovery of an amphibamid temnospondyl from the Early Permian of Texas that bridges the gap between other Palaeozoic amphibians and the earliest known salientians<sup>19,20</sup> and caudatans<sup>21</sup> from the Mesozoic. The presence of a mosaic of salientian and caudatan characters in this small fossil makes it a key taxon close to the batrachian (frog and salamander) divergence. Phylogenetic analysis suggests that the batrachian divergence occurred in the Middle Permian, rather than the late Carboniferous as recently estimated using molecular clocks<sup>18,22</sup>, but the divergence with caecilians corresponds to the deep split between temnospondyls and lepospondyls, which is congruent with the molecular estimates.

Tetrapoda Haworth, 1825

Temnospondyli Zittel, 1888

Amphibamidae Moodie, 1909

*Gerobatrachus hottoni* gen. et sp. nov.

## Holotype

United States National Museum of Natural History (Smithsonian Institute) (USNM) 489135. Discovered by P. Kroehler, a Museum Specialist at the USNM.

## Etymology

*Geros* (Greek), meaning aged or elder, and *batrachus* (Greek), meaning frog. Specific epithet is in honour of the late N. Hotton, vertebrate palaeontologist from the USNM.

## Locality and horizon

Locality number USNM 40971, 'Don's Dump Fish Quarry', Clear Fork Group, Baylor County, Texas, USGS Soap Creek 7.5' quad. More specific locality information is on file at the USNM.

## Age

Early Permian, Leonardian.

## Diagnosis

Amphibamid temnospondyl with 21 tiny pedicellate teeth on the premaxilla, and 17 presacral vertebrae; shares with crown group

salamanders a *basale commune* (combined distal tarsals 1 and 2) and tuberculum interglenoideum ('odontoid process') on atlas; shares with salientians and caudates an anteroposteriorly reduced vomer; shares with *Triadobatrachus* and crown group frogs a rod-like, laterally directed palatine; shares with *Karaurus*, *Triadobatrachus* and crown group frogs a broad skull, shortened presacral vertebral column; shares with most temnospondyls, frogs and basal salamanders a pedal phalangeal formula of ?-2-3-4-3; shares with frogs, *Amphibamus*, *Doleserpeton*, *Platyrhinops* and *Eoscopus* a large otic notch closely approaching the orbit; shares with frogs, salamanders, caecilians, *Amphibamus*, *Tersomius* and *Doleserpeton* pedicellate teeth; shares with *Amphibamus*, *Doleserpeton* and *Platyrhinops* a fore-shortened supratemporal; shares with *Amphibamus*, *Doleserpeton*, frogs and salamanders a foreshortened parasphenoid basal plate with wide lateral processes.

The holotype and only known specimen of *Gerobatrachus hottoni* was found in a two foot thick lens of fine-grained red siltstone sitting on the top of a knob, which was subsequently entirely excavated. The 110-mm-long specimen (Fig. 1) is preserved fully articulated in ventral view, and is missing only the stylopods, zeugopods, and ventral portions of the skull and pectoral girdle.

Most strikingly, the broad skull shape, the greatly enlarged vacuities on the palate, and the shortened vertebral column and tail give the immediate impression of a Palaeozoic batrachian. The premaxilla bears at least 21 small, pedicellate, monocuspid teeth that are not labiolingually compressed (Figs 2, 3a), a remarkable number for such a small element, and similar to the condition in batrachians. The frontals flare laterally at their anterior margin, as in derived amphibamids, and formed the dorsal orbital margin. The presence of a large parietal foramen near the frontoparietal suture indicates that this skeleton belonged to a juvenile individual (Fig. 2). The postparietals are surprisingly long elements in *Gerobatrachus*, but this unusual condition can be attributed to their exposure in internal view in this skull. Tabulars are restricted to the posterolateral corners of the skull table, and bear a hook-like posterior process, or 'horn', that extends posterior to the presumed location of the occiput.

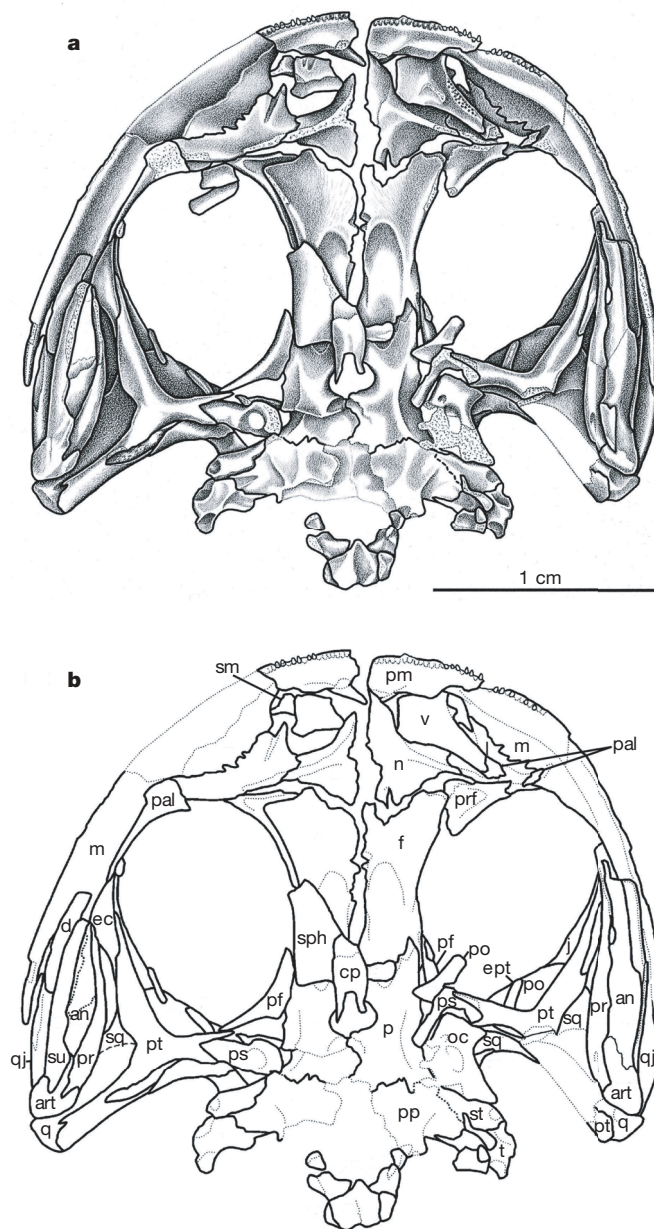
The palate and braincase are only partially preserved, but the exposed portions show several batrachian features. The vomer is anteroposteriorly narrow (not a broad plate as in other amphibamids), lacks palatal fangs, and has teeth restricted to a few rows on a raised patch along the medial margin of the choana. At its posterolateral extremity a portion of the rod-like, laterally directed palatine can be seen, a feature seen in *Triadobatrachus* and most crown group frogs<sup>20,23</sup>. Dorsal to the basicranial process of the salientian-like pterygoid, a small, rod-like, anterior projection is present, identical to epipterygoids described in the archaeobatrachian *Leiopelma*<sup>24</sup>. The pterygoid is prevented from reaching the lateral margin of the palate

<sup>1</sup>Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, 3330 Hospital Drive, Calgary, Alberta T2N 4N1, Canada.

<sup>2</sup>Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada. <sup>3</sup>Redpath Museum, McGill University, 859 Sherbrooke Street West, Montréal, Québec H3A 2K6, Canada. <sup>4</sup>Department of Biology, California State University at San Bernardino, 5500 University Parkway, San Bernardino, California 92407-2307, USA.

*Gerobatrachus* has 17 presacral vertebrae, which is transitional in number between other derived amphibamids (~21) and the salientian *Triadobatrachus* (14)<sup>20</sup> and caudatans *Karaurus* and *Chunerpeton* (14–15)<sup>21,25</sup>. As in salamanders, an anteriorly directed tuberculum intergenoideum of the atlas centrum is present, and at least the posterior vertebrae have narrow intercentra between holospondylous pleurocentra. Caudal vertebrae are very poorly ossified, similar to the condition seen in *Triadobatrachus* and some salamanders. The olecranon process of the ulna is surprisingly well-ossified for the inferred young ontogenetic stage of this specimen. The ilium lacks the posterior process common to temnospondyls but the presence of an anterior process, a salientian character, is obscured by an overlying fragment of the femur. An element identified as a sacral rib

Only two tarsal elements are present (Fig. 3b). A small, weakly ossified third distal tarsal is in articulation with the third metatarsal. At the base of the left first and second metatarsals is an elongate distal tarsal bone, broadly rounded distally but with a straighter margin proximally. Its position and large size is nearly identical with the combined distal tarsals 1 and 2, also called the *basale commune*, previously known exclusively in Caudata. While large enough to articulate with the proximal surfaces of metatarsals 1 and 2, it would



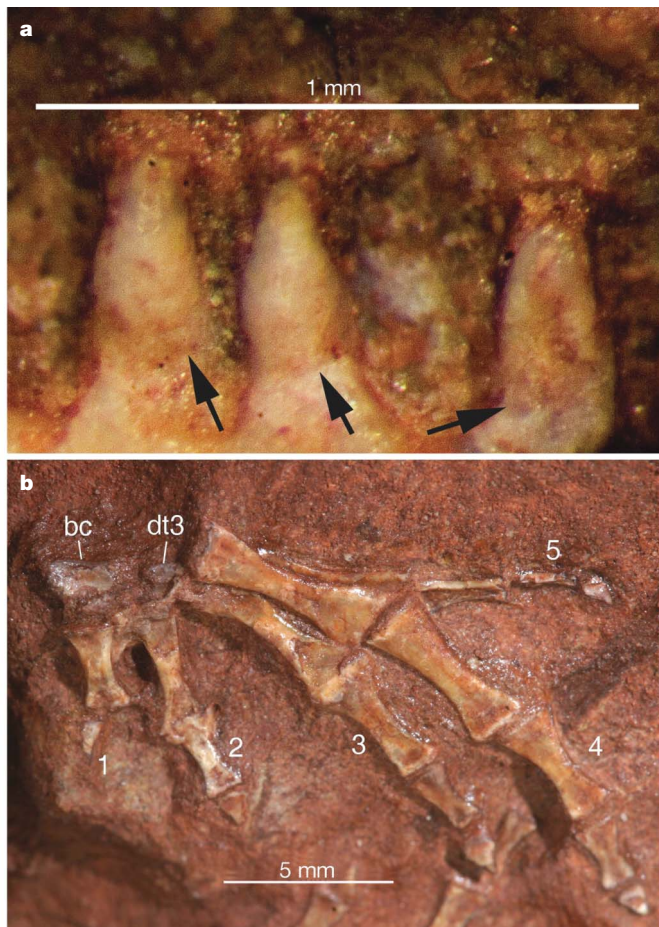
**Figure 2 | *Gerobatrachus hottoni*, gen. et sp. nov., holotype specimen USNM 489135.** **a**, Close-up interpretive specimen, and **b**, outline drawing of skull in ventral view. Abbreviations are the same as for Fig. 1 and: an, angular; art, articular; cp, cultriform process of parasphenoid; d, dentary; ec, ectopterygoid; ept, epipterygoid; f, frontal; j, jugal; l, lacrimal; m, maxilla; n, nasal; oc, portion of otic capsule; p, parietal; pal, palatine; pf, postfrontal; pm, premaxilla; po, postorbital; pp, postparietal; pr, prearticular; prf, prefrontal; ps, parasphenoid; pt, pterygoid; q, quadrate; jq, quadratojugal; sm, septomaxilla; sph, sphenethmoid; sq, squamosal; st, supratemporal; su, surangular; t, tabular; v, vomer.



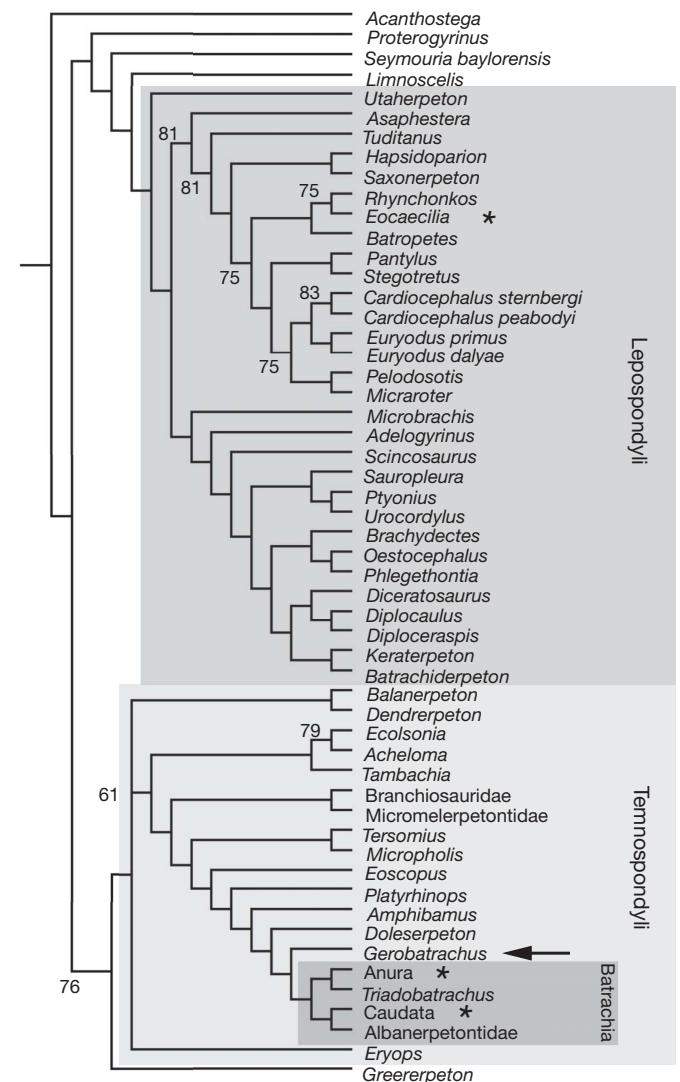
not do so completely, which is also the condition seen in salamanders. In extant salamanders the *basale commune* ossifies precociously<sup>26,27</sup>, a pattern that is consistent with this element being one of the only ossified tarsals in the juvenile skeleton of *Gerobatrachus*. Furthermore, the *basale commune* is the first mesopodial element to form during the initial mesenchymal condensation and chondrification and is a starting point for the establishment of the digital arch in a preaxial position, with subsequent condensations continuing post-axially<sup>27</sup>. This general directionality is mirrored by the subsequent ossification. Amniotes and frogs, on the contrary, ossify proximal mesopodial elements first, and then the distal postaxial elements, with the digital arch developing in a postaxial-to-preaxial direction. The presence of the *basale commune* and a more poorly ossified distal tarsal 3 as the only ossified mesopodial elements in *Gerobatrachus* suggests that it also may have had preaxial digital development. If our interpretations are correct, the preaxial pattern of digital development is either independently derived in *Gerobatrachus* and salamanders, or primitive in batrachians but reversed in frogs. Knowledge of development in fossil taxa is always inferential, especially when based on a single specimen, but our speculative hypothesis is testable with a more complete developmental series of either *Gerobatrachus* or another amphibamid. A preaxial pattern of digital development has recently been demonstrated in branchiosaurids<sup>28</sup>, which are thought to be closely related to, if not included within, Amphibamidae (Fig. 4), but branchiosaurids lack ossified carpals and tarsals and thus it remains unknown if they possessed a *basale*

*commune*. This observation, however, may support the possibility that preaxial development is primitive for batrachians (and more basal amphibamids), and will be the subject of future research.

We conducted a new phylogenetic analysis of basal tetrapod relationships to determine the placement of *Gerobatrachus* and test lissamphibian monophyly. A large matrix of lepospondyl relationships<sup>11</sup>, as recently modified<sup>12</sup>, was combined with a matrix of amphibamid relationships<sup>29</sup>. Duplicate characters were examined for inconsistencies in coding, which were rescored (based on direct observation of specimens whenever possible) if present, and then the duplicates were deleted. Redundant taxa were removed from the analysis. The number of taxa was further reduced to decrease computation time by eliminating highly fragmentary lepospondyl species. The final matrix (see Supplementary Information), containing 54 taxa and 219 characters, was subjected to parsimony analysis in PAUP\* 4.0b10. One hundred heuristic replicates (TBR branch swapping on shortest trees, random addition sequence) found 131 most parsimonious trees 1,125 steps long (consistency index 0.250, retention index 0.587; statistics calculated by PAUP\*).



**Figure 3 | *Gerobatrachus hottoni*, gen. et sp. nov., holotype specimen USNM 489135. a**, Close-up of left premaxillary teeth in lingual view, showing the presence of the dividing zone of poor ossification that separates the tooth cusp from the pedicel (indicated by arrows). **b**, Close-up photograph of the left pes, with the digital identification indicated by numbering. Abbreviations are the same as previous.



**Figure 4 | Majority rule consensus tree of 131 most parsimonious trees.** Numbers indicate the percentages of trees in which the given node appears, unnumbered nodes represent appearance in all trees. Lissamphibian taxa are indicated by \*, and *Gerobatrachus* is highlighted by an arrow. Recovery of lissamphibian monophyly within temnospondyls requires an additional 24–27 steps (Batrachia and Procera topologies, respectively), and recovery of lissamphibian monophyly within lepospondyls takes 30 additional steps.



Our analysis finds *Gerobatrachus* to be the immediate sister taxon to Batrachia (Fig. 4), with the amphibamids *Doleserpeton*, *Amphibamus* and *Platyrhinops* as successively more basal taxa. In addition, the oldest known caecilian *Eocaecilia* falls within recumbirostrine lepospondyls, sister group to *Rhynchonkos* and, one step further out, the brachystelechids. Thus, the available morphological evidence supports the hypothesis of a diphyletic origin of extant amphibians from Palaeozoic tetrapods, with a separate origin of the limbless, largely fossorial caecilians from within the lepospondyls, whereas Batrachia originates within Temnospondyli.

The discovery of a stem batrachian in the Early Permian places a new lower limit on the divergence between frogs and salamanders. *Gerobatrachus* is undeniably derived in comparison with other amphibamids, and therefore is most plausibly a recent addition to the Early Permian fauna, and not a relict form. The upper bound on the divergence is the occurrence of *Triadobatrachus* in the Triassic, so the divergence itself must have occurred between then and some point after the Early Permian, possibly the Middle Permian—(270–260) ± 0.7 Myr ago—considering the number of derived features *Gerobatrachus* shares with batrachians. Recent divergence estimates based on molecular clocks<sup>17,18</sup> are much older, placing this divergence in the late Carboniferous—308 ± 20 Myr ago<sup>18</sup>, and 357 ± 40 Myr ago<sup>17</sup>—although more recent unpublished estimates are much younger (D. San Mauro and D. Wake, personal communication). However, our finding of a diphyletic origin of lissamphibians places the divergence of batrachians and caecilians much earlier in tetrapod history, at the split between temnospondyls and lepospondyls. The minimum divergence of this event is 328–335 Myr ago, when the first temnospondyls and lepospondyls appear in the fossil record, which is much more consistent with the molecular estimates than implied by either of the monophyly hypotheses<sup>30</sup>.

Received 23 October 2007; accepted 25 February 2008.

- Milner, A. R. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) Vol. 1, *Amphibians, Reptiles, Birds* 59–102 (Clarendon Press, Oxford, 1988).
- Milner, A. R. The Paleozoic relatives of lissamphibians. *Herpetol. Monogr.* **7**, 8–27 (1993).
- Schoch, R. R. & Milner, A. R. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H. & Cloutier, R.) 345–377 (Verlag Dr. Friedrich Pfeil, München, 2004).
- Bolt, J. R. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 194–222 (Comstock Publishing Associates, Ithaca, 1991).
- Ruta, M. & Coates, M. I. Dates, nodes and character conflict: Addressing the lissamphibian origin problem. *J. Syst. Palaeontol.* **5**, 69–122 (2007).
- Ruta, M., Coates, M. I. & Quicke, D. L. Early tetrapod relationships revisited. *Biol. Rev.* **78**, 251–345 (2003).
- Trueb, L. & Cloutier, R. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 174–193 (Comstock Publishing Associates, Ithaca, 1991).
- Laurin, M. & Reisz, R. R. in *Amniote Origins* (eds Sumida, S. S. & Martin, K. L. M.) 9–59 (Academic, San Diego, 1997).
- Vallin, G. & Laurin, M. Cranial morphology and affinities of *Microbrachis*, and a reappraisal of the phylogeny and lifestyle of the first amphibians. *J. Vertebr. Paleontol.* **24**, 56–72 (2004).
- Laurin, M. The importance of global parsimony and historical bias in understanding tetrapod evolution. Part I. Systematics, middle ear evolution and jaw suspension. *Ann. Sci. Nat. Paris* **1**, 1–42 (1998).
- Anderson, J. S. The phylogenetic trunk: Maximal inclusion of taxa with missing data in an analysis of the Lepospondyli. *Syst. Biol.* **50**, 170–193 (2001).
- Anderson, J. S. in *Major Transitions in Vertebrate Evolution* (eds Anderson, J. S. & Sues, H.-D.) 182–227 (Indiana Univ. Press, Bloomington, 2007).
- Carroll, R. L. in *Amphibian Biology* (eds Heatwole, H. & Carroll, R. L.) Vol. 4, *Palaeontology: The Evolutionary History of Amphibians 1402–1411* (Surrey Beatty and Sons, Chipping Norton, Australia, 2000).
- Carroll, R. L. The importance of branchiosaurians in determining the ancestry of the modern amphibian orders. *Neues Jb. Geol. Palaeontol. Abh.* **232**, 157–180 (2004).
- Carroll, R. L. The Palaeozoic ancestry of salamanders, frogs and caecilians. *Zool. J. Linn. Soc.* **150** (s1), 1–140 (2007).
- Carroll, R. L. & Currie, P. J. Microsaurs as possible apodan ancestors. *Zool. J. Linn. Soc.* **57**, 229–247 (1975).
- San Mauro, D. et al. Initial diversification of living amphibians predated the breakup of Pangaea. *Am. Nat.* **165**, 590–599 (2005).
- Zhang, P. et al. Mitogenomic perspectives on the origin and phylogeny of living amphibians. *Syst. Biol.* **54**, 391–400 (2005).
- Piveteau, J. Paléontologie de Madagascar. XXIII. Un Amphibien du Trias inférieur. *Ann. Paléontol.* **26**, 135–177 (1937).
- Rage, J.-C. & Roček, Z. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the Early Triassic. *Palaeontographica A* **206**, 1–16 (1989).
- Ivachnenko, M. F. Urodelians from the Triassic and Jurassic of Soviet Central Asia. *Paleontol. J.* **12**, 362–368 (1978).
- San Mauro, D. et al. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. *Mol. Phylogenet. Evol.* **33**, 413–427 (2004).
- Duellman, W. E. & Trueb, L. *Biology of Amphibians* 2nd edn (The Johns Hopkins Univ. Press, Baltimore, 1994).
- Trueb, L. in *The Skull* (eds Hanken, J. & Hall, B. K.) Vol. 2, *Patterns of Structural and Systematic Diversity* 255–343 (Univ. Chicago Press, Chicago, 1993).
- Gao, K.-Q. & Shubin, N. H. Earliest known crown-group salamanders. *Nature* **422**, 424–428 (2003).
- Wang, Y. A new Mesozoic caudate (*Liaoxitriton dauhugouensis* sp. nov.) from Inner Mongolia, China. *Chin. Sci. Bull.* **49**, 858–860 (2004).
- Shubin, N. H. & Wake, D. B. in *Amphibian Biology* (eds Heatwole, H. & Davies, M.) Vol. 5, *Osteology 1782–1808* (Surrey Beatty and Sons, Chipping Norton, Australia, 2003).
- Fröbisch, N. B., Carroll, R. L. & Schoch, R. R. Limb ossification in the Paleozoic branchiosaurid *Apateon* (Temnospondyli) and the early evolution of preaxial dominance in tetrapod limb development. *Evol. Dev.* **9**, 69–75 (2007).
- Anderson, J. S. et al. *Georgenthalia clavinisica*, a new genus and species of dissorhoid temnospondyl from the Early Permian of Germany, and the relationships of the Family Amphibamidae. *J. Vertebr. Paleontol.* **28**, 61–75 (2008).
- Lee, M. S. Y. & Anderson, J. S. Molecular clocks and the origin(s) of modern amphibians. *Mol. Phylogenet. Evol.* **40**, 635–639 (2006).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Carrano, D. Chaney, B. DiMichele and P. Kroehler of the USNM for information and photographs of the discovery locality and for access to the specimen. E. Rega transported the specimen from Washington DC to Los Angeles. We thank P. Janvier and le Muséum National d'Histoire Naturelle in Paris for support while one of us (J.S.A.) studied the holotype of *Triadobatrachus*. The research was further supported by Natural Science and Engineering Research Council of Canada Discovery Grants to R.R.R. and J.S.A.

**Author Contributions** J.S.A. contributed to project planning, figure preparation, anatomical analysis, phylogenetic analysis, manuscript preparation and financial support for study; R.R.R. to phylogenetic analysis, manuscript preparation and financial support; D.S. to specimen preparation, figure preparation, anatomical analysis and manuscript preparation; N.B.F. to anatomical analysis, phylogenetic analysis and manuscript preparation; and S.S.S. to project initiation and manuscript preparation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.S.A. ([janders@ucalgary.ca](mailto:janders@ucalgary.ca)).

# The ground state of embryonic stem cell self-renewal

Qi-Long Ying<sup>1</sup>, Jason Wray<sup>2</sup>, Jennifer Nichols<sup>2</sup>, Laura Batlle-Morera<sup>2</sup>, Bradley Doble<sup>3</sup>, James Woodgett<sup>4</sup>, Philip Cohen<sup>5</sup> & Austin Smith<sup>2</sup>

**In the three decades since pluripotent mouse embryonic stem (ES) cells were first described<sup>1,2</sup> they have been derived and maintained by using various empirical combinations of feeder cells, conditioned media, cytokines, growth factors, hormones, fetal calf serum, and serum extracts<sup>1–7</sup>. Consequently ES-cell self-renewal is generally considered to be dependent on multifactorial stimulation of dedicated transcriptional circuitries, pre-eminent among which is the activation of STAT3 by cytokines (ref. 8). Here we show, however, that extrinsic stimuli are dispensable for the derivation, propagation and pluripotency of ES cells. Self-renewal is enabled by the elimination of differentiation-inducing signalling from mitogen-activated protein kinase. Additional inhibition of glycogen synthase kinase 3 consolidates biosynthetic capacity and suppresses residual differentiation. Complete bypass of cytokine signalling is confirmed by isolating ES cells genetically devoid of STAT3. These findings reveal that ES cells have an innate programme for self-replication that does not require extrinsic instruction. This property may account for their latent tumorigenicity. The delineation of minimal requirements for self-renewal now provides a defined platform for the precise description and dissection of the pluripotent state.**

Mouse ES cells exist in the artificial milieu of cell culture. They are derived and maintained by using a combination of the cytokine leukaemia inhibitory factor (LIF) to activate STAT3 and either serum or bone morphogenetic protein (BMP) to induce inhibitor-of-differentiation proteins<sup>5</sup>. Their differentiation involves autoinductive stimulation of the mitogen-activated protein kinase (ERK1/2) pathway by fibroblast growth factor-4 (FGF4)<sup>9,10</sup>. However, neither LIF nor serum/BMP block the activation of ERK (Supplementary Information and ref. 5). We proposed that the LIF and serum/BMP signals act downstream of phospho-ERK to block ES-cell commitment. To test this idea we used selective small-molecule inhibitors SU5402 (ref. 11) and PD184352 (ref. 12) to inhibit FGF receptor tyrosine kinases and the ERK cascade, respectively.

We found that, in combination with LIF, either inhibitor replaces the requirement for serum/BMP and supports robust long-term ES-cell propagation (Supplementary Information). Lineage commitment does not occur despite a reduced expression of inhibitor-of-differentiation proteins. In contrast, ES cells plated without LIF in either PD184352 or SU5402 progressively degenerate and cannot be maintained even though differentiation is suppressed. To reduce off-target side effects we tried low doses of PD184352 and SU5402 together (PS). In PS we find that undifferentiated ES cells expand through multiple passages (Fig. 1a, b). Differentiation is constrained, although occasional neural rosettes emerge. This result, observed with several independent ES cell lines, suggests that the minimal requirements for ES-cell self-renewal may be to deflect commitment signals emanating from FGF receptor and ERK signalling. However,

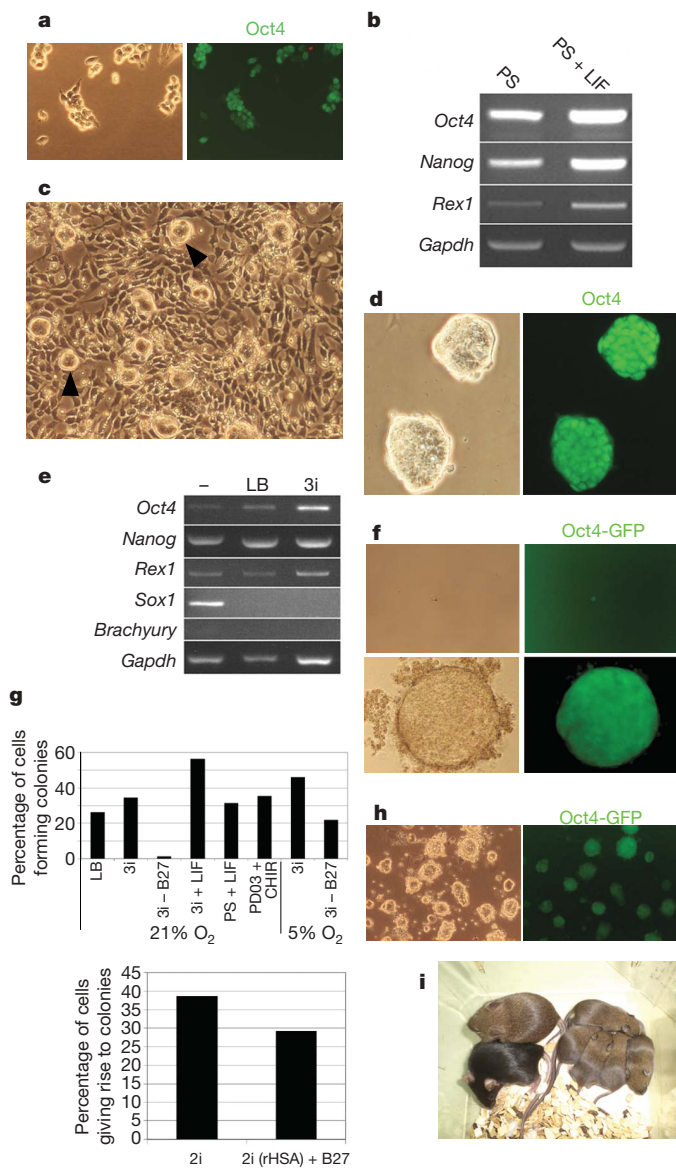
apoptosis is relatively high in PS, especially immediately after passage, and cells survive poorly at clonal density, which is indicative of collateral compromise to cell growth and viability.

ES-cell propagation has been reported to be enhanced by an indirubin entity, 6-bromo-indirubin-3'-oxime (BIO), that inhibits glycogen synthase kinase-3 (GSK3)<sup>4</sup>. However, indirubins are not highly selective and cross-react with cyclin-dependent kinases and other kinases<sup>13,14</sup>. We found reduced viability of ES cells in BIO with or without PS. Nevertheless we speculated that relief of GSK3-mediated negative regulation of biosynthetic pathways might restore growth to ES cells cultured in PS. We therefore used a more selective inhibitor, CHIR99021 (ref. 14,15). Alone, CHIR99021 enhances survival at low cell density but also induces non-neural differentiation. At higher densities some colonies remain morphologically undifferentiated but are progressively overcome by differentiation on passaging (Fig. 1c). Single blockade of GSK3 therefore has pleiotropic effects, promoting non-neural differentiation, suppressing neural differentiation and enhancing growth capacity. Crucially, however, in a combination of all three inhibitors (3i) the differentiation blocking effect of PS is dominant, resulting in a highly efficient expansion of undifferentiated colonies, even at a low cell density. Multiple ES-cell lines tested all expand continuously for many weeks in 3i. They express Oct4, Nanog and Rex1 with minimal levels of lineage commitment markers, Sox1 or brachyury (Fig. 1d, e). In 3i, ES cells expand with a doubling rate comparable to that in LIF plus serum/BMP (Supplementary Information) with the proportion of Oct4-green fluorescent protein (GFP)-positive undifferentiated cells remaining over 90%. As a rigorous test of the sufficiency of 3i to sustain ES-cell self-renewal, we examined the clonogenicity of isolated cells. After single-cell deposition, undifferentiated Oct4-positive colonies develop at higher frequency than in LIF and serum or BMP (Fig. 1f, g).

The B27 supplement used in serum-free culture contains defined additives, in particular antioxidants and free-radical scavengers. We found that ES cells could be propagated in bulk culture in 3i medium prepared with N2 supplement only, but they did not survive at clonal density. However, in physiological oxygen (5% O<sub>2</sub>) clonal propagation was obtained without B27 (Fig. 1g). This excludes an instructive contribution from B27 components to ES-cell self-renewal, while highlighting the damage potential of non-physiological oxygen levels. When insulin was omitted we found ES cells to be more sensitive to FGF receptor (FGFR) and MAP kinase/ERK kinase (MEK) inhibitors. We therefore decreased their concentrations. In these conditions, with only transferrin and albumin additives, ES cells expanded, even from single cells. They remained mostly undifferentiated over four weeks of continuous culture (Fig. 1h), although after the first passage the propagation rate declined steadily. We conclude that insulin promotes long-term growth capability but does not dictate the fate choice between self-renewal and lineage commitment.

<sup>1</sup>Center for Stem Cell and Regenerative Medicine, Department of Cell and Neurobiology, Keck School of Medicine, University of Southern California, 1501 San Pablo Street, ZNI 529, Los Angeles, California 90033, USA. <sup>2</sup>Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. <sup>3</sup>McMaster Stem Cell and Cancer Research Institute, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada. <sup>4</sup>Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. <sup>5</sup>Division of Signal Transduction Therapy and MRC Protein Phosphorylation Unit, University of Dundee, Dundee DD1 5EH, UK.

Finally, we used recombinant albumin to eliminate serum-derived contaminants. In combination with transferrin and insulin, this supported both bulk passing and clonal propagation (Fig. 1g).



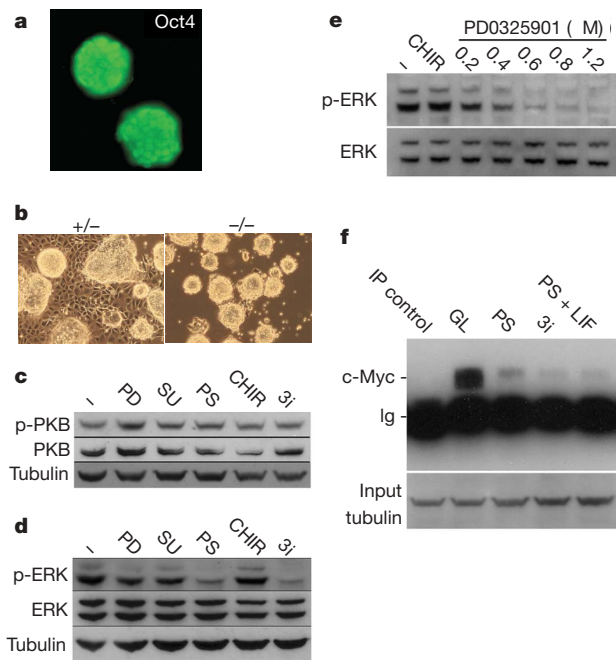
**Figure 1 | Three inhibitors (3i) support robust self-renewal and *de novo* derivation of pluripotent ES cells.** **a**, Immunostaining of E14Tg2a ES cells with Oct4 after four passages in N2B27 plus PD184352 and SU5402 (PS). **b**, RT-PCR analysis of marker expression in ES cells in N2B27 containing PS with or without LIF. *Gapdh*, gene encoding glyceraldehyde-3-phosphate dehydrogenase. **c**, Low-magnification phase-contrast image of ES cells passaged in N2B27 plus CHIR99021 showing a mixture of differentiated cells with compact undifferentiated colonies (arrowheads). **d**, Immunostaining with Oct4 after several passages in N2B27 plus 3i, showing compact colony morphology. **e**, RT-PCR analysis of marker expression in ES cells cultured in N2B27 alone (–) or with LIF and BMP4 (LB) or 3i. **f**, Phase and fluorescence images of expansion from a single Oct4GIP ES cell in 3i. **g**, Cloning efficiencies of E14Tg2a ES cells after single-cell deposition in the indicated conditions (top), and in CHIR99021 plus PD0325901 (2i, see Fig. 2) with or without B27, or with the replacement of serum albumin with recombinant albumin (rHSA) (bottom; experiment performed in 5% O<sub>2</sub>). **h**, Oct4GIP ES cells cultured for five passages (total 28 days) in basal medium supplemented with transferrin and BSA only plus 3 μM CHIR99021, 0.5 μM PD184352 and 1 μM SU5402. **i**, Chimaera and germline offspring produced from CBA ES cells derived in 3i. Chimaera showing extensive contribution of CBA (agouti coat colour) ES cells mated with C57BL/6 (black) produced agouti pups, indicating the transmission of the CBA genome.

To eliminate the possibility that self-renewal in 3i might reflect pre-adaptation to specific culture conditions in our laboratory, we investigated the derivation of ES cells from mouse embryos. ES cells were readily derived from blastocysts of the permissive 129 strain plated directly into 3i on gelatin-coated plastic. Expanded lines injected into blastocysts gave chimaeras and germline transmission (Supplementary Information). ES cell lines were also established from the CBA strain, which is refractory to ES-cell production under standard conditions<sup>16</sup>. Two of these lines were injected into morulae and both yielded high-grade chimaeras and germline transmission (Fig. 1i).

Taken together, the above findings demonstrate that 3i liberates ES cells from requirements for exogenous factors without compromise to developmental potency.

To confirm that blockade of FGF signalling is the critical target of SU5402 we substituted an alternative inhibitor, PD173074 (ref. 17). We found that this could substitute for SU5402 in 3i at 40-fold lower concentrations, which is consistent with its higher affinity for the FGF receptor (Fig. 2a). We then examined *fgf4*-null ES cells<sup>18</sup> and determined that they can expand continuously in CHIR99021 alone (Fig. 2b), providing genetic validation of the significance of autoinductive FGF4.

FGF4 activates the phosphatidylinositol-3-OH kinase/protein kinase B (PKB) and the Ras–MEK–ERK intracellular signalling cascades (Fig. 2c, d). Phosphorylation and activation of PKB is not appreciably altered by the 3i inhibitors. PD184352 or SU5402 applied alone at the low doses used in 3i cause only modest decreases in steady-state phospho-ERK. However, the combination of both



**Figure 2 | Effects of 3i components on intracellular signalling cascades.** **a**, E14Tg2a ES cells remain undifferentiated and Oct4-positive in alternative 3i with SU5402 replaced by PD173074. **b**, *fgf4*-null ES cells expand without differentiation in N2B27 plus CHIR99021 only, without a requirement for FGFR/MEK inhibition. **c**, **d**, Immunoblot analyses of steady-state levels of phospho(Ser 473)-PKB (p-PKB) (**c**) and phospho(Thr 202, Tyr 204)-ERK (p-ERK) (**d**) in ES cells after 24 h in N2B27 alone (–), plus 0.8 μM PD184352 (PD), 2 μM SU5402 (SU), 3 μM CHIR99021, PS or 3i. **e**, Immunoblot analyses of phospho(Thr 202, Tyr 204)-ERK levels in ES cells after 24 h in N2B27 alone (–), plus 3 μM CHIR99021 (CHIR) or 3 μM CHIR99021 plus PD0325901 at the indicated concentrations. **f**, c-Myc protein in ES cells assayed by sequential immunoprecipitation (IP) and immunoblotting after 24 h in serum plus LIF (GL), PS, 3i, or PS plus LIF. IP control is the GL sample immunoprecipitated with anti-tubulin. Input samples were subjected to SDS PAGE and blotted for tubulin to control for loading.



inhibitors greatly decreases phospho-ERK levels. CHIR99021 does not modulate phospho-ERK (Fig. 2e). We tested *erk2*-null ES cells<sup>19</sup> and found that these can be maintained at high density with CHIR99021 only, although optimal propagation requires supplementation with PD184352; this is consistent with maintained activity of phospho-ERK1 in these mutants. The central role of the ERK cascade was confirmed by using a structurally related, more potent but equally selective MEK inhibitor, PD0325901 (ref. 15), to achieve greater suppression of ERK activation without side effects. This is sufficient to sustain efficient ES-cell self-renewal in combination with CHIR99021 only (Figs 1g and 2e).

An unwarranted side effect of suppressing phospho-ERK is to depress *myc* messenger RNA and Myc protein levels (Fig. 2f and Supplementary Information). Upregulation of c-Myc has been suggested to mediate ES-cell self-renewal downstream of LIF and of BIO<sup>20</sup>. However, the low c-Myc levels in cultures in PS are not increased by CHIR99021 or LIF (Fig. 2f). Therefore elevated c-Myc is not necessary for ES-cell propagation, although some requirement for basal Myc activity is not excluded.

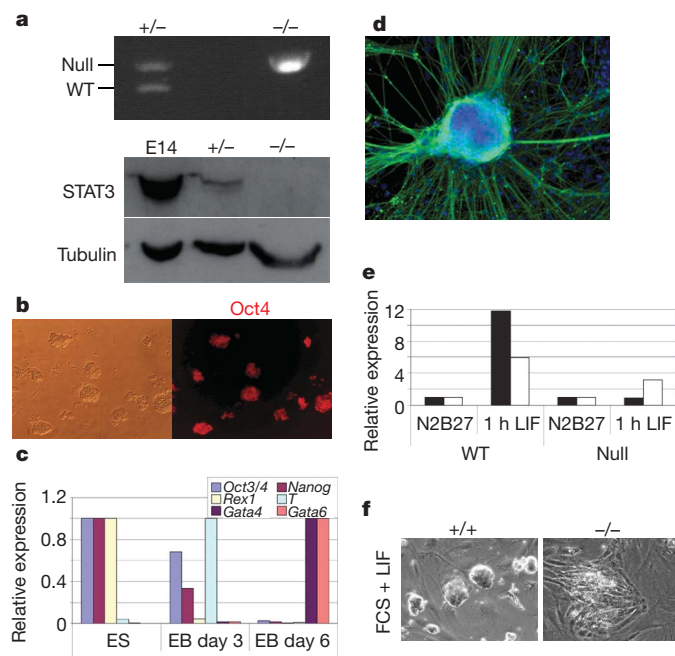
STAT3 signalling is central to previous models of ES-cell self-renewal<sup>8,21</sup> and has also been implicated in effects of BIO<sup>20,22,23</sup>. In 3i, however, we do not detect activation of STAT3 or induction of its target SOCS3 (Supplementary Information). To test definitively whether STAT3 is dispensable for ES-cell self-renewal, embryos from intercrosses of *Stat3* heterozygous mice were cultured in 3i. Homozygous mutant ES cells were established (Fig. 3a). *Stat3*-null cells are morphologically indistinguishable from wild-type ES cells. They express Oct4 and Nanog, and initiate multilineage commitment

in embryoid bodies (Fig. 3b–d). They show no induction of SOCS3 when stimulated with LIF (Fig. 3e). When transferred to LIF and serum, *stat3*<sup>-/-</sup> cells differentiate rapidly, confirming their incompetence to respond to LIF (Fig. 3f). We conclude that the otherwise absolute requirement for STAT3 in the derivation and self-renewal of mouse ES cells is rendered dispensable by 3i.

CHIR99021 induces a decrease in phosphorylation of  $\beta$ -catenin (Supplementary Information) and activation of the T-cell factor (TCF)-responsive TOPFlash reporter (Fig. 4a), simulating canonical Wnt signalling. We investigated whether Wnt could replicate the effect of CHIR99021. Recombinant Wnt3a alone induced non-neural differentiation, as seen with CHIR99021 only. This effect was suppressed by PS and at high concentrations (100 ng ml<sup>-1</sup>) Wnt3a seemed to eliminate residual neural differentiation and thereby improved ES-cell propagation. However, expansion in PS plus Wnt3a did not match that obtained in 3i. We introduced into ES cells dominant-negative  $\Delta$ NhLef1, which lacks the  $\beta$ -catenin-binding domain and suppresses TCF-mediated transcriptional activation. As expected, ES cells stably expressing  $\Delta$ NhLef1 showed reduced TOPFlash activity (Fig. 4a). Nonetheless they readily formed undifferentiated colonies in 3i. A competitive self-renewal assay was performed after treatment with Cre to excise the floxed  $\Delta$ NhLef1 and simultaneously activate GFP. Equivalent numbers of  $\Delta$ NhLef1-expressing and revertant GFP-expressing cells were propagated as mixed cultures for four passages. In serum plus LIF the GFP-positive and GFP-negative populations remained equivalent. In 3i the GFP-negative  $\Delta$ NhLef1-expressing cells became marginally predominant (Fig. 4b). Decreasing TCF activation therefore does not impede ES-cell self-renewal. Increased  $\beta$ -catenin levels might also enhance cell adhesion. However, E-cadherin-null ES cells that lack adhesion junctions remain undifferentiated and proliferate as rapidly in 3i as in LIF plus serum (Supplementary Information).

To confirm that the effect of CHIR99021 is mediated through the inhibition of GSK3, we interrogated ES cells in which both GSK3 $\alpha$  and GSK3 $\beta$  had been deleted<sup>24</sup>. These DKO cells are profoundly deficient in neural differentiation. They can be passaged two or three times in non-supplemented medium but succumb to progressive non-neural differentiation. This short-lived propagation is similar to that of wild-type ES cells cultured in CHIR99021 only (compare Fig. 4c with Fig. 2a). Addition of PS or PD0325901 eliminates differentiation and allows continuous passaging (Fig. 4c). However, expansion is slower than in wild-type cells in 3i. LIF restores normal population doubling (Fig. 4c), but CHIR99021 has no beneficial effect. This confirms that the effect of CHIR99021 is mediated through GSK3 and that LIF operates through a parallel STAT3 (refs 8, 21) pathway independent of GSK3 inhibition. DKO cells show constitutive TOPFlash activation<sup>24</sup>, 50-fold higher than CHIR99021-treated wild-type cells (Supplementary Information). This tonic  $\beta$ -catenin/TCF activity, with upregulation of targets such as brachyury and *cdx1* (ref. 24), probably underlies their compromised propagation.

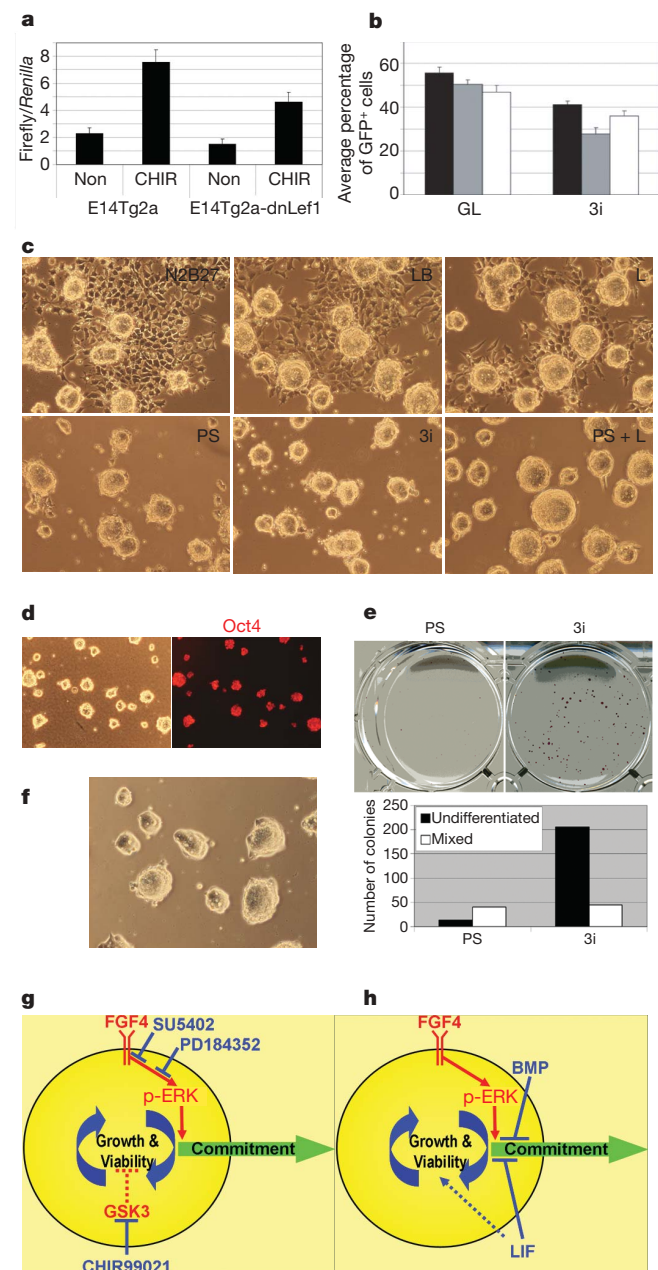
ES cells constitutively expressing elevated levels of Nanog are capable of sustained self-renewal in N2B27 alone but expand poorly at clonal density unless LIF is also added<sup>5</sup>. They form only a few small colonies at low density in PS but generate abundant undifferentiated colonies in 3i (Fig. 4e). The key effect of CHIR99021 therefore does not involve the induction of Nanog. Because Nanog-overexpressing ES cells are independently blocked in differentiation, this result further suggests that the contribution of GSK3 inhibition extends beyond limiting differentiation. To probe this further, we evaluated whether CHIR99021 could rescue ES cells subjected to a more profound blockade of phospho-ERK. A higher dose of PD0325901 (2 or 3  $\mu$ M) almost entirely eliminates phospho-ERK and causes growth arrest and cell death. The addition of CHIR99021 restores viability and allows efficient expansion of undifferentiated ES cells in the near absence of ERK signalling (Fig. 4f). We surmise that as phospho-ERK is diminished, downmodulation of GSK3 becomes increasingly



**Figure 3 | ES-cell propagation in 3i does not involve STAT3.** **a**, Top: genomic PCR for null and wild-type (WT) *Stat3* alleles in heterozygous and *Stat3* homozygous null ES cells. Bottom: immunoblot analysis of heterozygous and *Stat3* homozygous null ES cells. E14, E74Tg2a ES cells. **b**, Oct4 immunostaining of *Stat3*-null ES cells. **c**, Quantitative RT-PCR analysis of undifferentiated *Stat3*-null ES cells and derivative embryoid bodies (EB) at days 3 and 6. **d**, *Stat3*-null ES cells generate morphologically differentiated cells expressing the neuronal marker  $\beta$ III-tubulin (TuJ1 immunoreactive). **e**, Quantitative RT-PCR analysis of *Socs3* (*Stat3* target gene; filled columns) and *Egr1* (ERK target gene; open columns) expression in undifferentiated *Stat3* wild-type and null ES cells grown in N2B27 alone or stimulated with LIF for 1 h. **f**, *Stat3*-null MF1 ES cells differentiate in the presence of LIF and feeders, in contrast with wild-type MF1 ES cells, which remain undifferentiated.

crucial to maintain metabolic activity, biosynthetic capacity and overall viability.

This study reveals that the pathways required to sustain undifferentiated ES cells are dictated by the construction of the culture



**Figure 4 | CHIR99021 acts via inhibition of GSK3 to enhance ES-cell growth capacity and viability.** **a**, TOPFlash assay in  $\Delta$ NLef1 stable transfectants. Results are shown as means and s.d. for three biological replicates. Non, N2B27 alone; CHIR, CHIR99021. **b**, Competitive growth assay for three clones of  $\Delta$ NLef1 stable transfectants and Cre revertants (GFP<sup>+</sup>) in serum and LIF or in 3i. Results are shown as means and s.d. for four biological replicates. **c**, GSK3 $\alpha$ / $\beta$ -deficient DKO cells cultured in the indicated conditions for two passages. L, LIF. **d**, Phase-contrast and Oct4 immunostaining of DKO cells after eight passages in 1  $\mu$ M PD0325901. Parallel cultures with addition of CHIR99021 were indistinguishable. **e**, ES cells constitutively expressing Nanog respond to CHIR99021 by enhanced self-renewal at low density in 3i compared with PS. **f**, ES cells self-renew in 2  $\mu$ M PD0325901 plus CHIR99021. **g, h**, Diagrams of self-replication of the pluripotent state when inductive phospho-ERK signalling is either inhibited upstream by chemical antagonists (**g**) or counteracted downstream by LIF and BMP (**h**). Inhibition of GSK3 serves a key function in augmenting self-renewal when phospho-ERK (p-ERK) is suppressed by maintaining cellular growth capacity and additionally reinforcing suppression of neural commitment.

milieu. In a neutralized environment, ES cells can be efficiently derived and maintained without a requirement for growth factors or cytokines (Fig. 4g). We infer that BMP/Smad/Id and LIF/STAT3 signalling do not instruct self-renewal but act in unrefined culture conditions to shield the pluripotent state from induced phospho-ERK (Fig. 4h). Earlier studies have pointed to a positive effect of inhibiting the ERK cascade on ES-cell propagation in the context of additional signals<sup>25,26</sup>. However, upregulation of c-Myc, Stat3 or anti-apoptotic factors, previously invoked as key effectors of self-renewal, is not relevant in 3i. Our data do not exclude a contribution of stabilized  $\beta$ -catenin through TCF-independent mechanism(s), possibly acting as a noise filter<sup>27</sup>. Wnt3a does enhance neural suppression in PS cultures, but it gives significantly less benefit for overall propagation than CHIR99021 does. We infer that the pivotal contribution of GSK3 inhibition is to restore full growth and viability. This may be achieved by balancing the loss of ERK input into basic cellular processes. We detected no induction of anti-apoptotic factors (Supplementary Information), indicating that reduced GSK3 activity may exert a global modulation of the ES-cell metabolic and biosynthetic capacity rather than having a direct anti-apoptotic action. Furthermore, restoration of the biosynthetic capacity of ES cells might itself increase the threshold for commitment. This possibility is suggested by the effect of feedback in mitogen-activated protein kinase signalling circuitry on the mating switch decision in yeast<sup>28</sup>.

Previous empirical configurations of the culture environment have obscured the critical requirements for maintaining ES-cell pluripotency. We propose that ES cells are a basal cell state that is intrinsically self-maintaining if shielded effectively from inductive differentiation stimuli including autocrine FGF4. This feature may underlie the well-known predisposition of ES cells to generate teratocarcinomas<sup>29,30</sup>. They can dispense with an elementary cell signalling pathway, ERK, and do not seem to require any intercellular stimulation. They have not developed G1 cyclin checkpoint control of cell cycle progression and replicate constitutively<sup>29</sup>. ES cells thus display a self-sufficiency more akin to that of unicellular organisms than the interdependence generally exhibited by metazoan cells.

## METHODS SUMMARY

CHIR99021, PD184352 and PD0325901 were synthesized in the Division of Signal Transduction Therapy, University of Dundee. Inhibitors were used at the following concentrations unless otherwise specified: CHIR99021, 3  $\mu$ M; PD184352, 0.8  $\mu$ M; SU5402 (Calbiochem), 2  $\mu$ M; PD173074 (Sigma), 100 nM; PD0325901, 0.4  $\mu$ M in 3i or 1  $\mu$ M with CHIR99021 (2i). Clonal assays were performed by means of single-cell deposition into 96-well plates with a DakoCytomation MoFlo sorter.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 November 2007; accepted 2 April 2008.

- Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
- Evans, M. J. & Kaufman, M. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Rathjen, J. *et al.* Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J. Cell Sci.* **112**, 601–612 (1999).
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signalling by a pharmacological GSK-3-specific inhibitor. *Nature Med.* **10**, 55–63 (2004).
- Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
- Yoshida, K. *et al.* Maintenance of the pluripotent phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech. Dev.* **45**, 163–171 (1994).
- Ogawa, K., Matsui, H., Ohtsuka, S. & Niwa, H. A novel mechanism for regulating clonal propagation of mouse ES cells. *Genes Cells* **9**, 471–477 (2004).

8. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048–2060 (1998).
9. Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895–2902 (2007).
10. Stavridis, M. P., Lunn, J. S., Collins, B. J. & Storey, K. G. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* **134**, 2889–2894 (2007).
11. Mohammadi, M. *et al.* Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science* **276**, 955–960 (1997).
12. Davies, S. P., Reddy, H., Caivano, M. & Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.* **351**, 95–105 (2000).
13. Zhen, Y., Sorensen, V., Jin, Y., Suo, Z. & Wiedlocha, A. Indirubin-3'-monoxime inhibits autophosphorylation of FGFR1 and stimulates ERK1/2 activity via p38 MAPK. *Oncogene* **26**, 6372–6385 (2007).
14. Murray, J. T. *et al.* Exploitation of KESTREL to identify NDRG family members as physiological substrates for SGK1 and GSK3. *Biochem. J.* **384**, 477–488 (2004).
15. Bain, J. *et al.* The selectivity of protein kinase inhibitors; a further update. *Biochem. J.* **408**, 297–315 (2007).
16. Buehr, M. & Smith, A. Genesis of embryonic stem cells. *Phil. Trans. R. Soc. B* **358**, 1397–1402 (2003).
17. Mohammadi, M. *et al.* Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *EMBO J.* **17**, 5896–5904 (1998).
18. Wilder, P. J. *et al.* Inactivation of the FGF-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. *Dev. Biol.* **192**, 614–629 (1997).
19. Saba-El-Leil, M. K. *et al.* An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep.* **4**, 964–968 (2003).
20. Cartwright, P. *et al.* LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132**, 885–896 (2005).
21. Matsuda, T. *et al.* STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* **18**, 4261–4269 (1999).
22. Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochem. Biophys. Res. Commun.* **343**, 159–166 (2006).
23. Hao, J., Li, T. G., Qi, X., Zhao, D. F. & Zhao, G. Q. WNT/ $\beta$ -catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev. Biol.* **290**, 81–91 (2006).
24. Doble, B. W., Patel, S., Wood, G. A., Kockeritz, L. K. & Woodgett, J. R. Functional redundancy of GSK-3 $\alpha$  and GSK-3 $\beta$  in Wnt/ $\beta$ -catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev. Cell* **12**, 957–971 (2007).
25. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30–43 (1999).
26. Chen, S. *et al.* Self-renewal of embryonic stem cells by a small molecule. *Proc. Natl Acad. Sci. USA* **103**, 17266–17271 (2006).
27. Martinez-Arias, A. & Hayward, P. Filtering transcriptional noise during development: concepts and mechanisms. *Nature Rev. Genet.* **7**, 34–44 (2006).
28. Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).
29. Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.* **12**, 432–438 (2002).
30. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150–7160 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Alessi for discussion and advice on GSK3 signalling; A. Rizzino, S. Meloche and R. Kemler for *Fgf4*, *Erk2* and *Ecadherin* targeted ES cells, respectively; F. Watt for the  $\Delta$ NhLef1 construct; N. Shpiro and R. Marquez for synthesizing PD184352, CHIR99021 and PD0325901; B. Amati and G. Faga for advice on Myc immunoblotting; J. Vrana for fluorescence-activated cell sorting support; and C. Manson, K. Savill and colleagues for mouse husbandry. This research was funded by the Medical Research Council and the Biotechnology and Biological Sciences Research Council of the UK, the Canadian Institutes of Health Research, and by the European Commission Framework VI project EuroStemCell. P.C. is a Royal Society Research Professor, and A.S. is a Medical Research Council Professor.

**Author Contributions** Q.L.Y. and A.S. conceived the study; Q.L.Y., J.W. and J.N. designed, executed and interpreted experiments; L.B.M. generated CBA ES cells; B.D. and J.W. generated and provided GSK3 mutant ES cells; P.C. contributed expert advice and inhibitors; and A.S. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Q.L.Y. ([qying@keck.usc.edu](mailto:qying@keck.usc.edu)) or A.S. ([ags39@cscr.cam.ac.uk](mailto:ags39@cscr.cam.ac.uk)).



## METHODS

Details of RT-PCR conditions and antibodies are provided in Supplementary Information.

**ES-cell culture.** N2B27 medium was prepared as described<sup>31,32</sup> or with preformulated N2B27 base medium (catalogue no. SCS-SF-NB-02; Stem Cell Sciences Ltd.) Where specified, recombinant human albumin (Cellastim; Invitria) was substituted for bovine serum albumin. Cells were routinely propagated by trypsinization and replating every three days, with a split ratio of 1 in 10.

**ES-cell derivation.** Whole diapause blastocysts (strain 129) or isolated epiblasts (CBA) were plated in pre-equilibrated N2B27 3i medium. After several days, cell masses were disaggregated into small clumps of cells with trypsin and replated. Emerging ES-cell colonies were expanded by replating into successively larger wells. Wild-type MF1 ES cells were derived from whole blastocysts as for strain 129. These cells showed decreased substrate attachment, probably as a result of the outbred MF1 genetic background. They can be passaged on laminin-coated plastic but are more readily maintained on murine embryo fibroblast feeders. Accordingly, for *Stat3* mutant derivations we employed feeders for part of the process. Eight-cell embryos from intercrosses of *Stat3*<sup>+/-</sup> outbred MF1 mice<sup>33</sup> were cultured in KSOM medium containing 3i until the formation of expanded blastocysts. Trophectoderm layers were removed by immunosurgery and analysed by PCR for genotype determination<sup>34</sup>. Four isolated inner-cell masses (three *Stat3*<sup>+/-</sup>, one *Stat3*<sup>-/-</sup>) were plated individually into four-well plates on feeders in pre-equilibrated N2B27 3i. After five days, cell masses were disaggregated with trypsin and plated into fresh four-well plates. ES cells developed in all four cultures and were expanded in laminin-coated wells without feeders.

**Transfections.** ΔNhlLef1 (ref. 35) was inserted between loxP sites in pPyFloxMTIPgfp<sup>36</sup>. This construct was transfected by electroporation into E14Tg2a ES cells for stable integration, or into E14T ES cells<sup>36</sup> for episomal propagation. Dual luciferase reporter assays were performed 56 h after lipofection with TOPFlash or FOPFlash constructs.

**Competitive self-renewal assay.** Three clones of E14Tg2a stable transfectants were transiently transfected with a Cre expression vector to excise ΔNhlLef1 and simultaneously activate GFP expression. After 24 h, GFP-positive and GFP-negative cells were fractionated by fluorescence-activated cell sorting (FACS), recombined in equal numbers and plated in six-well plates at 10<sup>5</sup> cells per well in N2B27 with 3i. Cells were expanded for four passages and then analysed by FACS to establish the proportion of GFP<sup>+</sup> (ΔNhlLef1 excised) and GFP<sup>-</sup> (ΔNhlLef1-expressing) cells.

31. Brewer, G. J., Torricelli, J. R., Evege, E. K. & Price, P. J. Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. *J. Neurosci. Res.* **35**, 567–576 (1993).
32. Ying, Q. L. & Smith, A. G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
33. Takeda, K. *et al.* Targeted disruption of the mouse *Stat3* gene leads to early embryonic lethality. *Proc. Natl Acad. Sci. USA* **94**, 3801–3804 (1997).
34. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
35. Niemann, C., Owens, D. M., Hulsken, J., Birchmeier, W. & Watt, F. M. Expression of ΔNhlLef1 in mouse epidermis results in differentiation of hair follicles into squamous epidermal cysts and formation of skin tumours. *Development* **129**, 95–109 (2002).
36. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).

## LETTERS

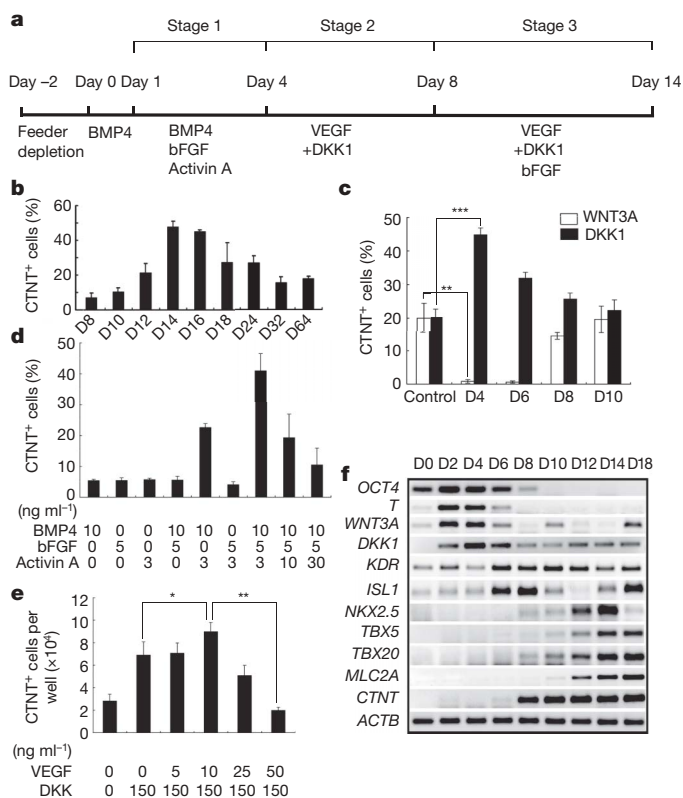
# Human cardiovascular progenitor cells develop from a KDR<sup>+</sup> embryonic-stem-cell-derived population

Lei Yang<sup>1</sup>, Mark H. Soonpaa<sup>2</sup>, Eric D. Adler<sup>1</sup>, Torsten K. Roepke<sup>3</sup>, Steven J. Kattman<sup>4</sup>, Marion Kennedy<sup>4</sup>, Els Henckaerts<sup>5</sup>, Kristina Bonham<sup>6</sup>, Geoffrey W. Abbott<sup>3</sup>, R. Michael Linden<sup>1,5</sup>, Loren J. Field<sup>2</sup> & Gordon M. Keller<sup>1,4</sup>

The functional heart is comprised of distinct mesoderm-derived lineages including cardiomyocytes, endothelial cells and vascular smooth muscle cells. Studies in the mouse embryo and the mouse embryonic stem cell differentiation model have provided evidence indicating that these three lineages develop from a common Flk-1<sup>+</sup> (kinase insert domain protein receptor, also known as Kdr) cardiovascular progenitor that represents one of the earliest stages in mesoderm specification to the cardiovascular lineages<sup>1</sup>. To determine whether a comparable progenitor is present during human cardiogenesis, we analysed the development of the cardiovascular lineages in human embryonic stem cell differentiation cultures. Here we show that after induction with combinations of activin A, bone morphogenetic protein 4 (BMP4), basic fibroblast growth factor (bFGF, also known as FGF2), vascular endothelial growth factor (VEGF, also known as VEGFA) and dickkopf homolog 1 (DKK1) in serum-free media, human embryonic-stem-cell-derived embryoid bodies generate a KDR<sup>low</sup>/C-KIT (CD117)<sup>neg</sup> population that displays cardiac, endothelial and vascular smooth muscle potential *in vitro* and, after transplantation, *in vivo*. When plated in monolayer cultures, these KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells differentiate to generate populations consisting of greater than 50% contracting cardiomyocytes. Populations derived from the KDR<sup>low</sup>/C-KIT<sup>neg</sup> fraction give rise to colonies that contain all three lineages when plated in methylcellulose cultures. Results from limiting dilution studies and cell-mixing experiments support the interpretation that these colonies are clones, indicating that they develop from a cardiovascular colony-forming cell. Together, these findings identify a human cardiovascular progenitor that defines one of the earliest stages of human cardiac development.

To direct the differentiation of human embryonic stem cells (ESCs) to the cardiac lineage, we designed a staged protocol that involved the formation of a primitive-streak-like population (stage 1, Fig. 1a), the induction and specification of cardiac mesoderm (stage 2) and the expansion of the cardiovascular lineages (stage 3) using combinations of factors known to have a role in these developmental steps in other systems<sup>2–7</sup>. Recent studies have shown that the combination of BMP4 and activin A will promote cardiac development in human ESC cultures<sup>8</sup>. However, the stage at which these pathways function in the establishment of this lineage was not defined. Using the protocol developed here, the combination of activin A and BMP4 at stage 1 induces a primitive-streak-like population and mesoderm, as demonstrated by the upregulation and transient expression of *T* (brachyury) and *WNT3A*—genes known to be expressed in these populations in the mouse<sup>9,10</sup> (Fig. 1f). At stage 2, the WNT inhibitor DKK1 is added to

specify cardiac mesoderm and VEGF is included to promote the expansion and maturation of the KDR<sup>+</sup> population. bFGF is added again at day 8 of differentiation to support the continued expansion of the



**Figure 1 | Specification of the cardiac lineage from human ESCs.** **a**, An outline of the protocol used for the differentiation of human ESCs to the cardiac lineage. **b**, Kinetics of CTNT<sup>+</sup> cell development in embryoid bodies induced with BMP4, bFGF and activin A. **c**, Frequency of CTNT<sup>+</sup> cells in day-14 embryoid bodies after manipulation of the WNT signalling pathway, as indicated. Control, cultures that did not receive WNT or DKK1. **d**, Frequency of CTNT<sup>+</sup> cells in day-14 embryoid bodies after induction with different combinations of BMP4, bFGF and activin A at stage 1. **e**, Effect of varying VEGF concentrations at stage 2 on the total number of CTNT<sup>+</sup> cells generated at day 14. **f**, Gene expression analysis of embryoid bodies at different stages of development. Where shown, bars represent standard error of the mean of three independent experiments; \**P* = 0.07, \*\**P* < 0.01, \*\*\**P* < 0.001. D0, D4, and so on, refer to days of culture.

<sup>1</sup>Department of Gene and Cell Medicine, The Black Family Stem Cell Institute, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, New York 10029, USA. <sup>2</sup>Wells Center for Pediatric Research, Indiana University School of Medicine, 1044 West Walnut Street, Indiana 46202, USA. <sup>3</sup>Greenberg Division of Cardiology, Departments of Medicine and Pharmacology, Weill Medical College of Cornell University, 520 East 70th Street, New York, New York 10021, USA. <sup>4</sup>McEwen Centre for Regenerative Medicine, University Health Network, 101 College Street, Toronto, Ontario M5G 1L7, Canada. <sup>5</sup>Department of Infectious Diseases, King's College London, London SE1 9RT, UK. <sup>6</sup>VistaGen Therapeutics Inc., 384 Oyster Point Boulevard, Suite 8, San Francisco, California 94080, USA.

developing cardiovascular lineages<sup>5</sup>. This protocol supports cardiac development, as demonstrated by the emergence of contracting embryoid bodies (Supplementary Fig. 1c) and cells that express cardiac troponin T (CTNT, also known as TNNT2; Fig. 1b),  $\alpha$ -actinin,  $\alpha/\beta$  myosin heavy chain, ANP (atrial natriuretic peptide) and connexin 43 (Supplementary Movie 1 and Supplementary Fig. 1a–b, d). The highest frequency of CTNT<sup>+</sup> cells is routinely observed between days 14 and 16 of culture (Fig. 1b and Supplementary Fig. 1b).

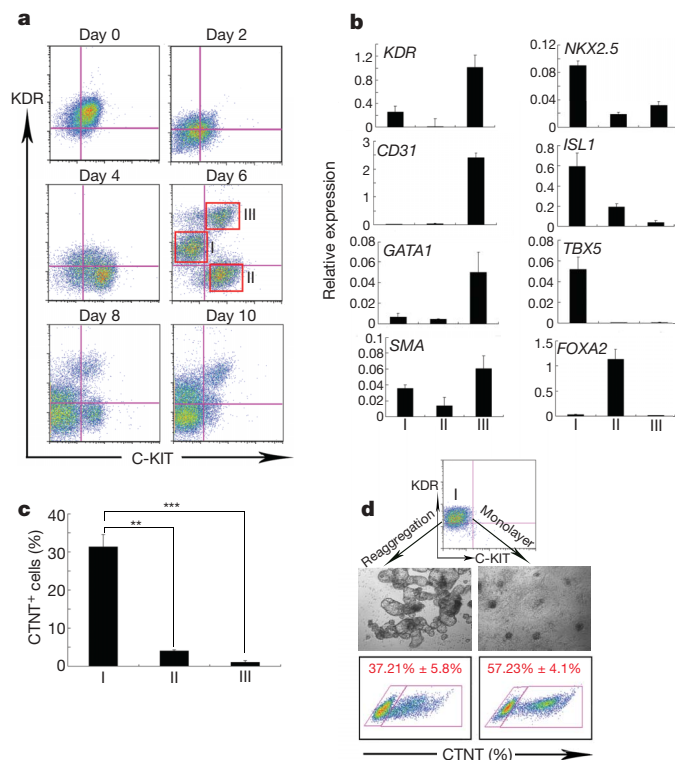
Because studies in other systems have shown that stage-specific inhibition of canonical WNT signalling is required for cardiac development<sup>4,6,11</sup>, we investigated the role of this pathway in the emergence of the cardiac lineage from the human ESCs, specifically focusing on stage 2. Addition of DKK1 at day 4 of differentiation led to a more than twofold increase in the frequency of CTNT<sup>+</sup> cells (up to 40%) at day 14 (Fig. 1c). The effect of DKK1 was less pronounced if added after day 4. WNT3A had the opposite effect and completely suppressed development of CTNT<sup>+</sup> cells if added at days 4 or 6. Taken together, these findings indicate that stage-specific inhibition of the canonical WNT pathway is necessary to promote cardiac specification of the BMP4/activin-A-induced primitive streak population. To evaluate the role of BMP4, activin A and bFGF, single factors as well as different combinations were tested during the

induction stage (stage 1). BMP4, bFGF or activin A alone or in combinations (BMP4 and FGF or activin A and bFGF) were poor inducers of cardiac development, as demonstrated by the low frequency (Fig. 1d) and low total number of CTNT<sup>+</sup> cells generated (Supplementary Fig. 2a, b). Although BMP4 and activin A did induce significant numbers of CTNT<sup>+</sup> cells, the combination of the three factors was the most potent and generated the highest frequency (40–50% CTNT<sup>+</sup>) and largest number of cardiac cells. Cardiac development was not dependent on exogenous VEGF. However, the addition of 10 ng ml<sup>-1</sup> of this factor did increase the total number of CTNT<sup>+</sup> cells generated (Fig. 1e and Supplementary Fig. 2c, d).

Molecular analysis of the developing embryoid bodies revealed dynamic changes in expression patterns after the establishment of the primitive-streak-like population. Together with *T* and *WNT3A*, expression of *DKK1* was upregulated early and persisted throughout the time course (Fig. 1f). *KDR* was expressed in undifferentiated ESCs. The levels of expression increased between days 4 and 6 and then persisted for the following 12 days. *ISL1*, a gene that marks progenitors of the secondary heart field in the early embryo<sup>12</sup>, was expressed between days 6 and 8, preceding the expression of the cardiac transcription factor *NKX2.5* (ref. 13), which was first detected at day 8 of differentiation. Expression of two TBX transcription factors required for cardiac development, *TBX5* (ref. 14) and *TBX20* (ref. 15), as well as the contractile proteins *MLC2A* (also known as *MYL7*) and *CTNT* was upregulated between days 8 and 10, reflecting the onset of cardiac development.

Recent studies with mouse ESC differentiation cultures identified a Flk-1<sup>+</sup> cardiovascular progenitor that develops from a Flk-1 population distinct from the hemangioblast<sup>1</sup>. To determine whether the human cardiac lineage also develops from a *KDR* (Flk-1<sup>+</sup>) population, we analysed developing embryoid bodies for expression of *KDR* and *C-KIT*. *C-KIT* was used because its expression in mouse embryoid bodies identifies the earliest hemangioblast-derived haematopoietic and vascular progenitors as well as the anterior primitive streak and the developing endoderm<sup>16</sup>. As shown in Fig. 2a, three distinct populations, *KDR*<sup>high</sup>/*C-KIT*<sup>+</sup> (III), *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> (I) and *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> (II), were detected at 6 days of differentiation. Development of the three populations was dependent on induction with both BMP4 and activin A (not shown). The *KDR*<sup>high</sup>/*C-KIT*<sup>+</sup> population expressed *CD31* (also known as *PECAM1*), VE-cadherin (*CDH5*) and *SMA* (smooth muscle actin), genes associated with vascular development, and *GATA1*, a gene indicative of haematopoietic commitment (Fig. 2b and Supplementary Fig. 3). Genes involved in cardiac development, including *NKX2.5*, *ISL1* and *TBX5*, were expressed at highest levels in the *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> fraction. This fraction also expressed *SMA*, very low levels of *GATA1*, but no detectable *CDH5* or *CD31*. The *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> cells expressed the highest levels of *OCT4*, *T*, *FOXA2* and *SOX17*, indicating the presence of residual undifferentiated ESCs and primitive-streak-like cells undergoing commitment to the endoderm lineage (Supplementary Fig. 3). *SOX1* and *PAX6* were detected at very low levels, suggesting little differentiation to the neuroectoderm lineage (Supplementary Fig. 3). Taken together, these expression patterns suggest that the *KDR*<sup>high</sup>/*C-KIT*<sup>+</sup> population contains haematopoietic and vascular progenitors, that the *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> population contains cardiac progenitors and that the *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> population consists of undifferentiated ESCs, primitive-streak-like cells and endodermal cells.

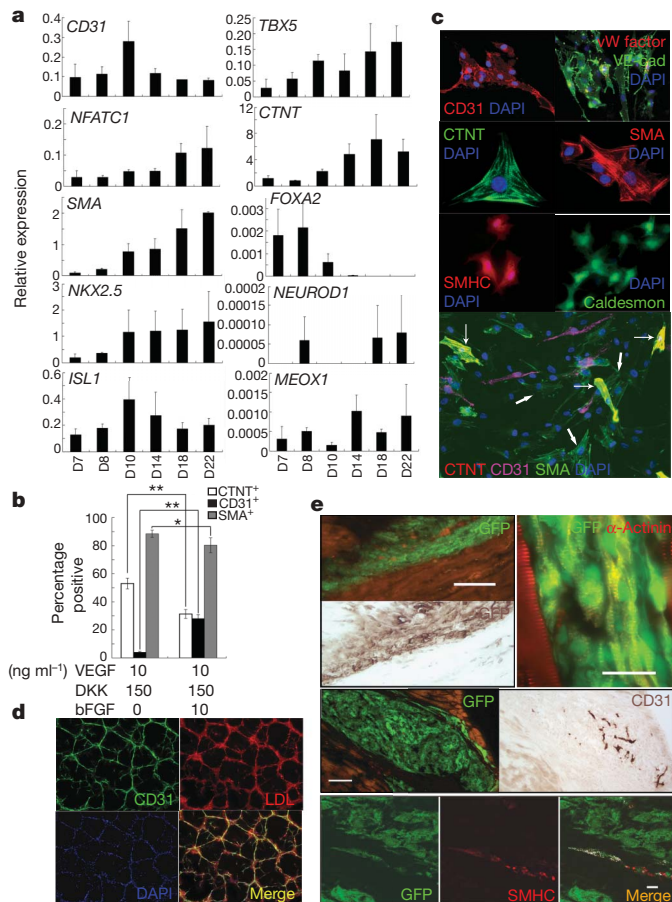
Consistent with the above expression profile, the *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> population displayed the greatest cardiomyocyte potential (Fig. 2c) and readily generated CTNT<sup>+</sup> cells and populations of contracting cells when cultured either as aggregates in suspension cultures or as adherent monolayers (Fig. 2d and Supplementary Movies 2 and 3). Approximately 40% of the aggregates and more than 50% of the monolayers were CTNT<sup>+</sup> after 7–10 days of culture, reflecting efficient differentiation to the cardiac lineage (Fig. 2d). The high frequency of cardiomyocytes in the monolayer cultures



**Figure 2 | Identification and characterization of the cardiovascular *KDR*<sup>low</sup>/*C-KIT*<sup>neg</sup> embryoid body population.** **a**, Flow cytometric analysis of different aged embryoid bodies, demonstrating the development of the three distinct populations (*KDR*<sup>high</sup>/*C-KIT*<sup>+</sup> (III), *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> (I) and *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> (II)) defined by co-expression of *KDR* and *C-KIT*. **b**, Quantitative RT-PCR gene expression analysis of the *KDR*<sup>high</sup>/*C-KIT*<sup>+</sup>, *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> and *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> populations isolated from day-6 embryoid bodies. The average expression normalized to cyclophilin is shown. **c**, Frequency of CTNT<sup>+</sup> cells generated from the *KDR*<sup>high</sup>/*C-KIT*<sup>+</sup>, *KDR*<sup>low</sup>/*C-KIT*<sup>+</sup> and *KDR*<sup>neg</sup>/*C-KIT*<sup>+</sup> populations cultured as monolayers in the presence of VEGF (10 ng ml<sup>-1</sup>), DKK1 (150 ng ml<sup>-1</sup>) and bFGF (10 ng ml<sup>-1</sup>). Cells were assayed after 10 days of culture. **d**, Cardiac potential of *KDR*<sup>low</sup>/*C-KIT*<sup>neg</sup> cells from day-6 embryoid bodies cultured as a monolayer or as aggregates in low-cluster wells with VEGF (10 ng ml<sup>-1</sup>) and DKK1 (150 ng ml<sup>-1</sup>). CTNT<sup>+</sup> cells were analysed after 7–10 days of culture. Where shown, bars represent standard error of the mean of three independent experiments; \*\**P* < 0.01, \*\*\**P* < 0.001.



routinely led to the development of sheets of cells contracting as a synchronous mass (Supplementary Movie 3). The isolated  $KDR^{low}/C-KIT^{neg}$  cells expanded approximately 1.5-fold as aggregates (data not shown) and 3-fold in the monolayer cultures (Supplementary Fig. 4b). With this induction protocol, we estimate an output of one



**Figure 3 | Characterization of the  $KDR^{low}/C-KIT^{neg}$ -derived lineages.** **a**, Quantitative RT-PCR analysis of adherent populations generated from the day-6  $KDR^{low}/C-KIT^{neg}$  fraction cultured with VEGF ( $10 \text{ ng ml}^{-1}$ ), DKK1 ( $150 \text{ ng ml}^{-1}$ ) and bFGF ( $10 \text{ ng ml}^{-1}$ ). (D7 represents one day following plating). The average expression normalized to cyclophilin is shown. **b**, Effect of bFGF on differentiation of day-6  $KDR^{low}/C-KIT^{neg}$  cells. Cells were harvested and analysed by flow cytometry after 10–12 days of culture. **c**, Immunostaining analysis of the day-6  $KDR^{low}/C-KIT^{neg}$ -derived population cultured in the presence of VEGF ( $10 \text{ ng ml}^{-1}$ ), DKK1 ( $150 \text{ ng ml}^{-1}$ ) and bFGF ( $10 \text{ ng ml}^{-1}$ ). Thin white arrows in the bottom panel indicate cardiac cells that express both CTNT and SMA; thick white arrows mark putative VSM cells that express SMA but not CTNT. Magnification;  $\times 400$  (CTNT, SMA),  $\times 200$  (all other panels). vW factor, von Willebrand factor. **d**, CD31 immunostaining and Dil-AC-LDL uptake of  $KDR^{low}/C-KIT^{neg}$ -derived cells cultured on Matrigel-coated glass coverslips. Magnification,  $\times 50$ . **e**, Derivation of cardiac lineages from the  $KDR^{low}/C-KIT^{neg}$  population in the mouse heart. Top left, fluorescent micrograph of grafted heart viewed with a FITC/TRITC (fluorescein isothiocyanate/tetramethylrhodamine isothiocyanate) cube, demonstrating the presence of GFP-positive donor cells (green, upper panel). Below this is a light micrograph of immuno-histochemical staining of same section with anti-GFP antibody visualized with DAB (brown signal, lower panel). Scale bar,  $100 \mu\text{m}$ . Top right, immunostaining of donor-derived GFP cells with anti- $\alpha$ -actinin antibody (yellow). Scale bars,  $10 \mu\text{m}$ . Middle left, donor-derived GFP cells. Middle right, immunohistochemical staining of serial section with human-specific anti-pecan (CD31) antibody (brown signal). Scale bar,  $100 \mu\text{m}$ . Bottom, confocal images of grafted heart stained with anti-smooth muscle MHC antibody (red). Green signal indicates GFP-expressing grafted cells in the same section. Colocalization is indicated with the white signal. Scale bar,  $10 \mu\text{m}$ . Where shown, bars represent standard error of the mean of three independent experiments; \* $P = 0.099$ , \*\* $P < 0.01$ .

cardiomyocyte per four input human ESCs. Kinetic analysis of embryoid bodies generated from another human ESC line (H1) demonstrated the development of the three  $KDR/C-KIT$  populations at day 5 rather than day 6 of differentiation. Analysis of the day-5  $KDR^{low}/C-KIT^{neg}$  population indicated that it also displays cardiac potential (Supplementary Fig. 5).

Expression analysis (quantitative PCR, qPCR) of the  $KDR^{low}/C-KIT^{neg}$ -derived adherent populations at different days after plating demonstrated the upregulation of genes associated with endothelial (CD31, *CDH5*), VSM (vascular smooth muscle) (calponin, *SMA*), cardiac development (*NKX2.5*, *ISL1*, *TBX5*, *TBX20*) and cardiac maturation (*CTNT*, *MLC2A*) (Fig. 3a and Supplementary Fig. 4a). Expression of *NFATC1* and neuregulin 1 (*NRG1*) suggests the presence of endocardium in the cultures<sup>17,18</sup>. The low levels of *NEUROD1*, *PAX6*, *SOX1*, *FOXA2*, *FOXA3*, *SOX17* and *MEOX1* expression indicate little, if any, contamination of these cultures with neuroectoderm, endoderm or somitic mesoderm (Fig. 3a and Supplementary Fig. 4a). Flow cytometric analysis of the  $KDR^{low}/C-KIT^{neg}$ -derived adherent population cultured for 10–12 days in VEGF and DKK1 revealed that almost 90% expressed SMA, 50% expressed CTNT and 4% expressed CD31 (Fig. 3b). Addition of bFGF to the cultures reduced the proportion of CTNT<sup>+</sup> and SMA<sup>+</sup> cells to 30% and 80%, respectively, and increased the CD31<sup>+</sup> subpopulation to 30%. The addition of bFGF did not significantly influence cell numbers in the monolayer cultures (Supplementary Fig. 4b). These findings indicate that the majority of cells within the  $KDR^{low}/C-KIT^{neg}$ -derived population are of the cardiovascular lineages and that bFGF can influence the proportion of cardiomyocytes and endothelial cells that develop in this population.

Immunostaining analysis of the  $KDR^{low}/C-KIT^{neg}$ -derived population demonstrated the presence of CD31<sup>+</sup>, VE-cadherin<sup>+</sup> and von Willebrand factor<sup>+</sup> endothelial cells, of CTNT<sup>+</sup> cardiomyocytes, and of SMA<sup>+</sup>, SMHC<sup>+</sup> (also known as MYH11<sup>+</sup>) and caldesmon<sup>+</sup> cells, indicative of VSM development (Fig. 3c). The immature cardiomyocytes within the population expressed both CTNT and SMA (lower panel of Fig. 3c, thin arrows), whereas the VSM cells expressed only SMA (thick arrows).  $KDR^{low}/C-KIT^{neg}$ -derived cells that were expanded in the presence of VEGF and bFGF formed a lattice, indicative of the formation of tube-like structures when cultured on Matrigel-coated coverslips. The cells within these structures expressed CD31 and displayed the capacity to take up Dil-AC-LDL (1,1'-dioctadecyl-1,3,3,3',3'-tetramethylindocarbocyanine perchlorate acetylated low-density lipoprotein), confirming their endothelial phenotype (Fig. 3d). The findings from the immunostaining analysis are consistent with those from the flow cytometric studies, and demonstrate that the  $KDR^{low}/C-KIT^{neg}$ -derived population consists of cells of the cardiac, endothelial and vascular smooth muscle lineages.

$KDR^{low}/C-KIT^{neg}$ -derived cells generated from a green fluorescent protein (GFP)-expressing hES2 cell line (GFP-hES2) were transplanted into the hearts of non-obese diabetic/severe combined immunodeficient (NOD/SCID) mice to document their developmental potential *in vivo*. Histological analyses revealed the presence of GFP<sup>+</sup> donor cells detected by epifluorescence and by staining with an anti-GFP antibody (Fig. 3e, top left panel). GFP<sup>+</sup> populations co-expressing  $\alpha$ -actinin (Fig. 3e, top right panel), CD31 (Fig. 3e, middle panel) or SMHC (Fig. 3e, lower panel) were detected in the grafts, indicating differentiation to the cardiac, endothelial and vascular smooth muscle lineages *in vivo*, respectively. Teratomas were not observed in any of the transplanted animals.  $KDR^{low}/C-KIT^{neg}$ -derived cells were also transplanted directly into infarcted hearts of SCID beige mice. When analysed two weeks later, animals transplanted with the  $KDR^{low}/C-KIT^{neg}$ -derived cardiovascular population had a 31% higher ejection fraction than those injected with media alone ( $56\% \pm 3.6\%$  versus  $39\% \pm 4.8\%$  (mean  $\pm$  s.e.m.,  $P = 0.008$ )). These findings are consistent with previous reports<sup>8,19,20</sup> in demonstrating that transplantation of human ESC-derived

cardiomyocytes leads to improvement in cardiac function in rodent models of myocardial infarction. Although such cell transplantation does improve function, it is important to stress that the mechanisms mediating this effect are currently not known<sup>21</sup>.

To establish the lineage relationship between these three cell types, we adapted the methylcellulose colony assay used to identify the cardiovascular progenitor in mouse ESCs cultures<sup>1</sup>. When plated in methylcellulose,  $KDR^{low}/C-KIT^{neg}$ -derived cells generated small compact colonies within 4 days of culture (Fig. 4a, light). PCR analysis of individual 4-day-old colonies demonstrated co-expression of

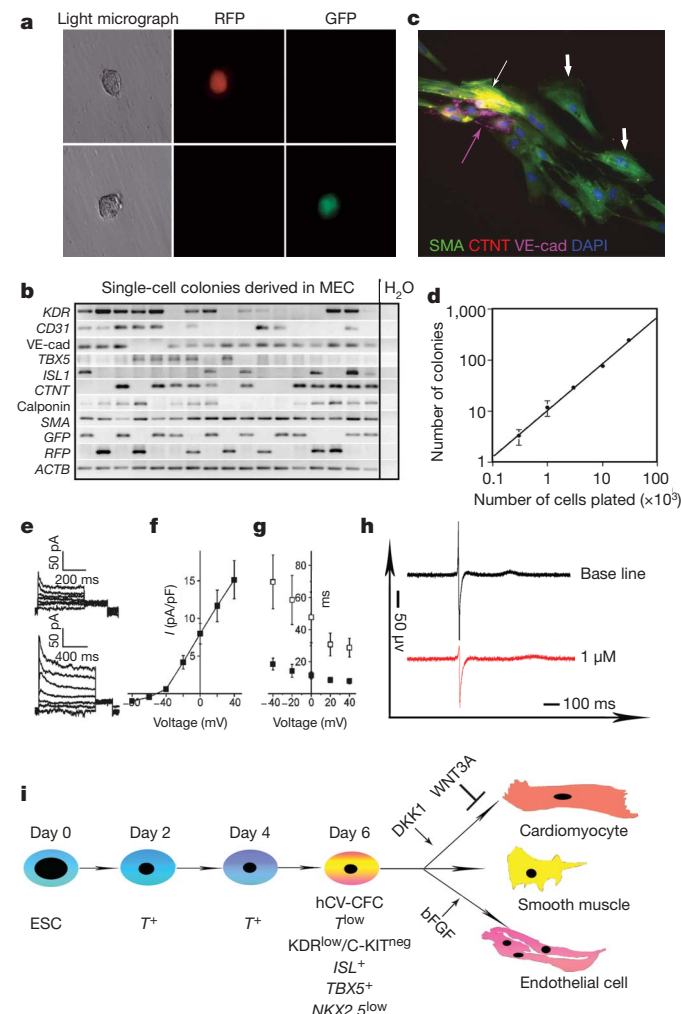
markers indicative of cardiac (*CTNT*), endothelial (*CD31* and/or *CDH5*) and VSM (*SMA* and/or calponin) development (Fig. 4b and Supplementary Fig. 4c). When maintained in culture for a further 6 days, a portion of these colonies generated contracting cardiomyocytes (Supplementary Movie 4). *ISL1* and *TBX5* were typically not expressed in the same colonies, suggesting that their expression may define colonies that contain distinct cardiac subpopulations from different heart fields. Immunostaining of adherent populations from individual colonies confirmed the presence of the cardiac, endothelial and VSM lineages (Fig. 4c).

Two different approaches were used to determine if the cardiovascular colonies are clonal. First,  $KDR^{low}/C-KIT^{neg}$ -derived cells from the GFP-hES2 cell line and from a human ESC line expressing red fluorescent protein (RFP) (HES2.R26)<sup>22</sup> were mixed in the methylcellulose assay. Colonies that developed expressed either GFP or RFP, but not both (Fig. 4a, b), consistent with the interpretation that they arise from a single cell and not from cell aggregation. In the second approach, we carried out a cell-dose response experiment. The relationship between the number of colonies that developed and the number of cells plated was linear, with a slope approaching one, further supporting the notion that the colonies are derived from a single cell (Fig. 4d). Taken together, these findings strongly suggest that these colonies represent clones of cardiovascular cells derived from a cardiovascular colony-forming cell (hCV-CFC).

The functional potential of  $KDR^{low}/C-KIT^{neg}$ -derived cardiomyocytes was evaluated with whole-cell current and field potential measurements. In whole-cell voltage-clamp analysis, 80% of cells studied expressed a predominant voltage-gated, transient outward potassium current (Fig. 4e, f). The voltage dependence, density and gating kinetics of this current (Fig. 4f, g) resembled those of the  $I_{to}$  (transient outward  $K^+$  current) potassium current found in human atrial and ventricular myocytes<sup>23,24</sup>. Field potential recordings using microelectrodes revealed that the  $KDR^{low}/C-KIT^{neg}$ -derived cardiac cells were electrically coupled to one another. In addition, as expected, the Vaughan Williams class-Ia-agent quinidine decreased the measured T-wave amplitude and increased the QT interval<sup>25</sup> (Fig. 4h).

Recent studies in the mouse have provided evidence that the main lineages of the heart develop from a common cardiovascular progenitor<sup>1,26,27</sup>. The identification of a hCV-CFC in this study suggests that cardiovascular development in humans is similar to this and that the cardiac, endothelial and vascular smooth muscle lineages are derived from a common progenitor (model, Fig. 4i). A recent study has reported the identification of a  $C-KIT^+KDR^-$  cardiac stem cell in the adult human heart that also displays the capacity to generate myocytes, smooth muscle cells and endothelial cells<sup>28</sup>. The differences in surface markers between the two populations may reflect the fact that one represents an early embryonic stage of development whereas the other is derived from the adult heart.

In mouse ESC cultures, the cardiovascular progenitor develops from a Flk-1 population that is distinct from the Flk-1 population containing the hemangioblast. Although we have not measured the temporal development of the two  $KDR$  populations in this study, we have previously demonstrated that the first haematopoietic progenitors to develop in human embryoid bodies are present in a  $KDR^+/C-KIT^+$  fraction, comparable to the  $KDR^{high}/C-KIT^+$  fraction in this study<sup>29</sup>. The finding that the cardiovascular progenitors are detected in a distinct  $KDR^{low}/C-KIT^{neg}$  fraction demonstrates a similar segregation of early haematopoietic and cardiac potential in the human system. The identification of the  $KDR^{low}/C-KIT^{neg}$  population that contains cardiovascular progenitors provides a unique opportunity to investigate the mechanisms that regulate the onset of human cardiac development as well as those that control their specification to the cardiac and vascular lineages. Access to this population also provides an enriched source of progenitors for engineering cardiovascular tissue *in vitro* and for transplantation



**Figure 4 | Identification and characterization of human cardiovascular progenitors.** **a**, Four-day-old cardiovascular colonies from mixed cultures showing of either RFP or GFP expression. **b**, Expression analysis of 4-day-old cardiovascular colonies isolated from the mixed RFP/GFP cultures. **c**, Immunostaining of cells grown from a single colony. The thin white arrow indicates cardiac cells that express both CTNT and SMA; the thick white arrows mark putative VSM cells that express SMA but no CTNT. The thin purple arrow identifies VE-cadherin<sup>+</sup> endothelial cells. Magnification,  $\times 200$ . **d**, Cell-dose response showing the relationship between the number of  $KDR^{low}/C-KIT^{neg}$ -derived cells plated and the number of cardiovascular colonies that develop. Error bars, s.e.m. **e**, Exemplar traces showing whole-cell voltage-clamp recordings of transient outward  $K^+$  current ( $I_{to}$ ) natively expressed in  $KDR^{low}/C-KIT^{neg}$ -derived cardiomyocytes. **f**, Mean current density–voltage relationship for cells as in **e**. From a batch of ten cells, eight showed the  $I_{to}$  current; the mean  $\pm$  s.e.m. current densities were plotted using traces from these eight cells. **g**, Mean time-to-peak current (solid squares) and inactivation  $\tau$  (open squares) for cells as in **f** ( $n = 8$ ). **h**, Extracellular electrical activity recorded without (base line) or with  $1 \mu M$  quinidine (lower red line) from cells cultured in a MEA (multi-channel system) chamber. **i**, Model depicting development of cardiovascular progenitors (hCV-CFC) in human ESC cultures.

to large animal models that may accurately reflect human cardiac function.

## METHODS SUMMARY

**Maintenance and differentiation of human ESCs.** The different human ESC lines were maintained as described<sup>29</sup>. For differentiation to the cardiac lineage, the following cytokines were used: days 0–1, BMP4 (0.5 ng ml<sup>-1</sup>); days 1–4, BMP4 (10 ng ml<sup>-1</sup>), bFGF (5 ng ml<sup>-1</sup>) and activin A (3 ng ml<sup>-1</sup>); days 4–8, DKK1 (150 ng ml<sup>-1</sup>) and VEGF (10 ng ml<sup>-1</sup>); after day 8, VEGF (10 ng ml<sup>-1</sup>), DKK1 (150 ng ml<sup>-1</sup>) and bFGF (5 ng ml<sup>-1</sup>). Cultures were maintained in a 5% CO<sub>2</sub>/5% O<sub>2</sub>/90% N<sub>2</sub> environment for the first 10–12 days and were then transferred into a 5% CO<sub>2</sub>/air environment.

**In vitro analysis of the KDR<sup>low</sup>/C-KIT<sup>neg</sup> population.** Isolated KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells were cultured as a monolayer on a gelatin-coated surface or as aggregates in low-cluster wells in Stempro34 medium supplemented with VEGF (10 ng ml<sup>-1</sup>) and DKK1 (150 ng ml<sup>-1</sup>). Where indicated, bFGF (10 ng ml<sup>-1</sup>) was added to the cultures to promote endothelial development. For analysis of the endothelial lineage, KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells were cultured in Stempro34 medium supplemented with VEGF (25 ng ml<sup>-1</sup>) and bFGF (25 ng ml<sup>-1</sup>) for 5–7 days and then seeded onto Matrigel-coated glass coverslips for an additional 24 h.

**In vivo analyses of KDR<sup>low</sup>/C-KIT<sup>neg</sup>-derived populations.** KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells (100,000) derived from GFP–human ESCs were injected into the left ventricular wall of NOD/SCID- $\gamma$  mice. Hearts were harvested 2–11 weeks post surgery. Immunohistochemistry was carried out with anti-GFP antibody (Chemicon; Vector ABC and DAB kits),  $\alpha$ -actinin antibody (Sigma), CD31 antibody (Dako) and SM-MHC antibody (Biomedical Technologies). Confocal images were analysed for colocalization using ImageJ and Pierre Bourdoncle's plugin with default settings. Myocardial infarction was induced in SCID beige mice. After 10–20 min, the mice were injected with 500,000 KDR<sup>low</sup>/C-KIT<sup>neg</sup>-derived cells ( $n = 9$ ) or an equivalent volume of basic media ( $n = 12$ ). Two weeks later, assessment of ventricular function was performed using 9.4 tesla magnetic resonance imaging.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 24 January; accepted 22 February 2008.

Published online 23 April 2008.

- Kattman, S. J., Huber, T. L. & Keller, G. M. Multipotent Flk-1<sup>+</sup> cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* **11**, 723–732 (2006).
- Conlon, F. L. et al. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919–1928 (1994).
- Lough, J. et al. Combined BMP-2 and FGF-4, but neither factor alone, induces cardiogenesis in non-precordial embryonic mesoderm. *Dev. Biol.* **178**, 198–202 (1996).
- Marvin, M. J., Di Rocco, G., Gardiner, A., Bush, S. M. & Lassar, A. B. Inhibition of Wnt activity induces heart formation from posterior mesoderm. *Genes Dev.* **15**, 316–327 (2001).
- Mima, T., Ueno, H., Fischman, D. A., Williams, L. T. & Mikawa, T. Fibroblast growth factor receptor is required for *in vivo* cardiac myocyte proliferation at early embryonic stages of heart development. *Proc. Natl Acad. Sci. USA* **92**, 467–471 (1995).
- Schneider, V. A. & Mercola, M. Wnt antagonism initiates cardiogenesis in *Xenopus laevis*. *Genes Dev.* **15**, 304–315 (2001).
- Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105–2116 (1995).
- Laflamme, M. A. et al. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).

- Kispert, A. & Herrmann, B. G. Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev. Biol.* **161**, 179–193 (1994).
- Liu, P. et al. Requirement for *Wnt3* in vertebrate axis formation. *Nature Genet.* **22**, 361–365 (1999).
- Ueno, S. et al. Biphasic role for Wnt/ $\beta$ -catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 9685–9690 (2007).
- Cai, C. L. et al. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5**, 877–889 (2003).
- Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I. & Harvey, R. P. *Nkx-2.5*: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119**, 419–431 (1993).
- Bruneau, B. G. et al. Chamber-specific cardiac expression of *Tbx5* and heart defects in Holt–Oram syndrome. *Dev. Biol.* **211**, 100–108 (1999).
- Cai, C. L. et al. T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* **132**, 2475–2487 (2005).
- Gouon-Evans, V. et al. BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nature Biotechnol.* **24**, 1402–1411 (2006).
- de la Pompa, J. L. et al. Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. *Nature* **392**, 182–186 (1998).
- Meyer, D. et al. Isoform-specific expression and function of neuregulin. *Development* **124**, 3575–3586 (1997).
- Leor, J. et al. Human embryonic stem cell transplantation to repair the infarcted myocardium. *Heart* **93**, 1278–1284 (2007).
- Van Laake, L. W. et al. Human embryonic stem cell-derived cardiomyocytes survive and mature in the mouse heart and transiently improve function after myocardial infarction. *Stem Cell Res.* **1**, 9–24 (2007).
- Rubart, M. & Field, L. J. ES cells for troubled hearts. *Nature Biotechnol.* **25**, 993–994 (2007).
- Irion, S. et al. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nature Biotechnol.* **12**, 1477–1482 (2007).
- Shibata, E. F. et al. Contributions of a transient outward current to repolarization in human atrium. *Am. J. Physiol.* **257**, 1773–1781 (1989).
- Konarzewska, H., Peeters, G. A. & Sanguinetti, M. C. Repolarizing K<sup>+</sup> currents in nonfailing human hearts. Similarities between right septal subendocardial and left subepicardial ventricular myocytes. *Circulation* **92**, 1179–1187 (1995).
- Meyer, T. Micro-electrode arrays in cardiac safety pharmacology. *Drug Safety* **11**, 763–772 (2004).
- Moretti, A. et al. Multipotent embryonic *Isl1*<sup>+</sup> progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151–1165 (2006).
- Wu, S. M. et al. Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127**, 1137–1150 (2006).
- Bearzi, C. et al. Human cardiac stem cells. *Proc. Natl Acad. Sci. USA* **104**, 14068–14073 (2007).
- Kennedy, M. et al. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood* **109**, 2679–2687 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Oza for MEA (multi-electrode arrays) measurement and members of the Keller laboratory for critically reading this manuscript. G.M.K., S.J.K. and G.W.A. are supported by the National Institutes of Health/National Heart Lung and Blood Institute.

**Author Contributions** L.Y. carried out most of the experiments; L.Y., S.J.K. and G.M.K. designed the study; L.Y. and G.M.K. analysed the data and wrote the manuscript; M.H.S. and L.J.F. performed the transplantation and differentiation study in the normal hearts; E.D.A. was responsible for the transplantation and analyses of the infarcted hearts; T.K.R. and G.W.A. carried out the patch-clamp study; E.H. and R.M.L. generated the AAV (adeno-associated virus)–GFP–hES2 cells; M.K. provided advice on experimental design and analysed data; and L.Y. and K.B. performed the field potential recording.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.M.K. ([gkeller@uhnresearch.ca](mailto:gkeller@uhnresearch.ca)).



## METHODS

**Maintenance of human ESCs.** H1 ESCs (National Institutes of Health, NIH code WA01) from the WiCell Research Institute and hES2 ESCs (NIH code ES02) from ES Cell International (ESI) were maintained as described<sup>29</sup>. The AAVS1 (adeno-associated virus integration site 1)-targeted hES2 cell line was generated by co-infection of parental hES2 cells with  $10^6$  viral particles of both AAV2-TRUF11 (CAG-GFP-TK-neo) and wild-type AAV2. After G418 selection to deplete the cells not infected by AAV2-TRUF11, GFP-positive cells were sorted and subclones were isolated. Targeted integration of the transgenes was confirmed by ligation-mediated PCR. Wild-type AAV sequences were not detected in GFP-positive clones. Generation of the RFP-expressing hES2 cell line was previously described<sup>22</sup>.

**Differentiation of human ESCs.** Embryoid bodies for differentiation were generated as described previously<sup>29</sup>. In brief, embryoid bodies were formed by plating small aggregates of human ESCs in 2 ml basic media (StemPro34, Invitrogen, containing 2 mM glutamine,  $4 \times 10^{-4}$  M monothioglycerol (MTG),  $50 \mu\text{g ml}^{-1}$  ascorbic acid, Sigma, and  $0.5 \text{ ng ml}^{-1}$  BMP4). The following concentrations of factors were used for primitive-streak formation and for mesoderm induction and cardiac specification: BMP4,  $10 \text{ ng ml}^{-1}$ ; human bFGF,  $5 \text{ ng ml}^{-1}$ ; activin A,  $3 \text{ ng ml}^{-1}$ ; human DKK1,  $150 \text{ ng ml}^{-1}$ ; and human VEGF,  $10 \text{ ng ml}^{-1}$ . The factors were added with the following sequence: days 1–4, BMP4, bFGF and activin A; days 4–8, VEGF and DKK1; after day 8, VEGF, DKK1 and bFGF. All factors were purchased from R&D Systems. Cultures were maintained in a 5%  $\text{CO}_2$ /5%  $\text{O}_2$ /90%  $\text{N}_2$  environment for the first 10–12 days and then transferred to a 5%  $\text{CO}_2$ /air environment.

**Flow cytometry.** Embryoid bodies were harvested and dissociated to single cells with trypsin ( $0.25\%$  trypsin-EDTA). For intracellular FACS, cells were fixed and stained with primary and secondary antibodies in PBS plus  $0.5\%$  saponin (Sigma). Analyses were carried out using a FacsCalibur flow cytometer (Becton Dickinson). Cells were sorted from day-6 embryoid bodies using a MoFlo (Dako Cytomation) cell sorter. Data were analysed using the FlowJo (Treestar) software. Anti-KDR-PE and anti-C-KIT-APC (allophycocyanin) were purchased from R&D Systems.

**In vitro analysis of the KDR<sup>low</sup>/C-KIT<sup>neg</sup> population.** For cardiac differentiation, isolated KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells were cultured as a monolayer on a gelatin-coated surface or as aggregates in low-cluster wells in StemPro34 medium supplemented with VEGF ( $10 \text{ ng ml}^{-1}$ ) and DKK1 ( $150 \text{ ng ml}^{-1}$ ). Cells were seeded at a concentration of 40,000–50,000 per well in a 96-well plate. To promote endothelial development, bFGF ( $10 \text{ ng ml}^{-1}$ ) was added. For analysis of the endothelial lineage, KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells were cultured in StemPro34 medium with VEGF ( $25 \text{ ng ml}^{-1}$ ) and bFGF ( $25 \text{ ng ml}^{-1}$ ) for 5–7 days. To induce the formation of tube-like structures, the cells were transferred and cultured on Matrigel-coated glass coverslips for 24 h.

**Immunofluorescence.** Dissociated cells were cultured on glass cover slips for 2 days, fixed with 4% PFA for 20 min, and then stained. Cells were incubated with the primary antibody for 2 h at  $37^\circ\text{C}$ , washed three times and then incubated with a secondary antibody for an additional 1 h. After staining, the cells were washed three times, and fluorescence was visualized with a Leica DMRA2 fluorescence microscope (Leica) and recorded using a digital Hamamatsu CCD camera. The following antibodies were used for staining: anti-human CD31 and anti-human VE-cadherin from R&D Systems, anti-mouse troponin T and anti-human smooth muscle actin from Lab Vision, anti-human ANP, anti-connexin 43 and anti-human  $\alpha/\beta$  MHC from Chemicon, anti-human  $\alpha$ -actinin from Sigma, and anti-human smooth muscle myosin heavy chain (SMHC), anti-human caldesmon and anti-von Willebrand factor from DakoCytomation.

The Cy2-, Cy3- and Cy5-conjugated secondary antibodies were purchased from Jackson ImmunoResearch.

**Colony assays.** KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells isolated from day-6 embryoid bodies were aggregated in the presence of VEGF ( $25 \text{ ng ml}^{-1}$ ), bFGF ( $10 \text{ ng ml}^{-1}$ ) and DKK1 ( $150 \text{ ng ml}^{-1}$ ) for 2–3 days. The aggregates were dissociated with trypsin and the cells were cultured in methylcellulose containing VEGF ( $25 \text{ ng ml}^{-1}$ ), bFGF ( $25 \text{ ng ml}^{-1}$ ) and DKK1 ( $150 \text{ ng ml}^{-1}$ ) in a 5%  $\text{CO}_2$ /5%  $\text{O}_2$ /90%  $\text{N}_2$  environment. Colonies were scored after 4–6 days of culture.

**RT-PCR.** For expression studies, individual colonies were isolated from the methylcellulose cultures and analysed using a modified version of the protocol described<sup>30</sup>. The amplified complementary DNA was then subjected to normal PCR. Real-time quantitative PCR was performed on a MasterCycler EP RealPlex (Eppendorf). Experiments were done in triplicate using Platinum SYBR GreenER qPCR SuperMix (Invitrogen). All primers are described in Supplementary Table 1. All annealing reactions were carried out at  $60^\circ\text{C}$ .

**In vivo analyses of KDR<sup>low</sup>/C-KIT<sup>neg</sup>-derived populations.** Day-6 GFP<sup>+</sup> KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells cultured in the presence of VEGF ( $10 \text{ ng ml}^{-1}$ ), bFGF ( $10 \text{ ng ml}^{-1}$ ) and DKK1 ( $150 \text{ ng ml}^{-1}$ ) for 5–10 days were injected (100,000 per recipient) directly into the left ventricular wall of NOD/SCID-gamma mice in an open-chest procedure. Hearts were harvested 2–11 weeks post surgery, fixed in 1% PFA in cacodylate buffer and sectioned at  $10 \mu\text{m}$  ( $n = 15$ ). Immuno-histochemistry was done with GFP (Chemicon, AB3080; Vector ABC and DAB kits),  $\alpha$ -actinin (Sigma, A781), CD31 (Dako, M0823), and smooth muscle MHC antibodies (Biomedical Technologies, BT-562). Confocal images were analysed for colocalization using ImageJ and Pierre Bourdoncle's plugin with default settings. For evaluation in the infarct model, myocardial infarction was induced in SCID beige mice by means of direct coronary ligation. 1–20 min later the mice were injected with 500,000 KDR<sup>low</sup>/C-KIT<sup>neg</sup>-derived cells ( $n = 9$ ) or an equivalent volume of media ( $n = 12$ ). Two weeks later, assessment of ventricular function was performed using 9.4 Tesla magnetic resonance imaging.

**Patch clamp.** Whole-cell patch-clamp recordings were performed using an IX50 inverted microscope (Olympus), a Multiclamp 700A amplifier, a Digidata 1300 analogue/digital converter and a PC with pClamp9.1 software (Axon Instruments). The bath solution contained (in mM): NaCl 136, KCl 4,  $\text{CaCl}_2$  1,  $\text{MgCl}_2$  2,  $\text{CoCl}_2$  5, HEPES 10, glucose 10 and tetrodotoxin (TTX) 0.02 (pH 7.4). Pipettes were of 3–5 M $\Omega$  resistance when filled with intracellular solution containing (in mM): KCl 135, EGTA 10, HEPES 10 and glucose 5 (pH 7.2). Cells were stepped from a holding potential of  $-80 \text{ mV}$  to test potentials from  $-80 \text{ mV}$  to  $+40 \text{ mV}$  in  $20 \text{ mV}$  increments, before a  $-30 \text{ mV}$  tail pulse (durations as in Fig. 4f). Data were analysed using pClamp9.1 software (Axon Instruments). Current amplitudes were normalized to cell size (whole-cell membrane capacitance). Inactivation  $\tau$  values were calculated using a single exponential fit of current decay.

**Field potential recording.** KDR<sup>low</sup>/C-KIT<sup>neg</sup> cells were cultured in a MEA (Multi Channel Systems) chamber in StemPro34 with  $10 \text{ ng ml}^{-1}$  VEGF and  $150 \text{ ng ml}^{-1}$  DKK1 for 2–4 weeks. Two days before measuring recordings, the media was changed to DMEM (Mediatech) with 15% FBS. Extracellular electrical activity was simultaneously recorded from 60 channels and analysed with the software MC Rack (Multi Channel Systems).

**Data analysis.** Data are shown as mean  $\pm$  standard error of three independent experiments. Statistic analysis was performed with the Student's  $t$  test.

30. Brady, G. & Iscove, N. N. Construction of cDNA libraries from single cells. *Methods Enzymol.* 225, 611–623 (1993).

# Multi-genetic events collaboratively contribute to *Pten*-null leukaemia stem-cell formation

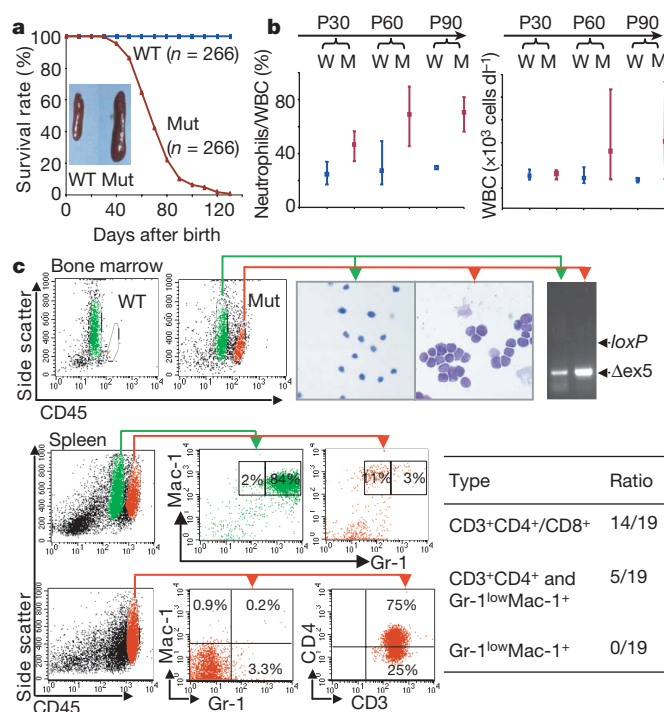
Wei Guo<sup>1</sup>, Joseph L. Lasky<sup>2</sup>, Chun-Ju Chang<sup>1</sup>, Sherly Mosessian<sup>1</sup>, Xiaoman Lewis<sup>1</sup>, Yun Xiao<sup>5</sup>, Jennifer E. Yeh<sup>6</sup>, James Y. Chen<sup>1</sup>, M. Luisa Iruela-Arispe<sup>3</sup>, Marileila Varella-Garcia<sup>5</sup> & Hong Wu<sup>1,4</sup>

Cancer stem cells, which share many common properties and regulatory machineries with normal stem cells, have recently been proposed to be responsible for tumorigenesis and to contribute to cancer resistance<sup>1</sup>. The main challenges in cancer biology are to identify cancer stem cells and to define the molecular events required for transforming normal cells to cancer stem cells. Here we show that *Pten* deletion in mouse haematopoietic stem cells leads to a myeloproliferative disorder, followed by acute T-lymphoblastic leukaemia (T-ALL). Self-renewable leukaemia stem cells (LSCs) are enriched in the c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>-</sup> compartment, where unphosphorylated  $\beta$ -catenin is significantly increased. Conditional ablation of one allele of the  $\beta$ -catenin gene substantially decreases the incidence and delays the occurrence of T-ALL caused by *Pten* loss, indicating that activation of the  $\beta$ -catenin pathway may contribute to the formation or expansion of the LSC population. Moreover, a recurring chromosomal translocation, T(14;15), results in aberrant overexpression of the *c-myc* oncogene in c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>-</sup> LSCs and CD3<sup>+</sup> leukaemic blasts, recapitulating a subset of human T-ALL. No alterations in Notch1 signalling are detected in this model, suggesting that *Pten* inactivation and *c-myc* overexpression may substitute functionally for Notch1 abnormalities<sup>2,3</sup>, leading to T-ALL development. Our study indicates that multiple genetic or molecular alterations contribute cooperatively to LSC transformation.

The PTEN–phosphatidylinositol-3-OH kinase (PI(3)K) pathway has been implicated in human leukaemogenesis<sup>2,4</sup>. Although acute deletion of the murine *Pten* gene in adult haematopoietic stem cells (HSCs) leads to defects in HSC self-renewal, it also causes a brief myeloproliferative disorder (MPD), followed by the development of acute leukaemia<sup>5,6</sup>. However, LSCs responsible for this transplantable disease have not been identified and, more significantly, the molecular mechanisms responsible for LSC formation remain to be elucidated.

Here we report the establishment of a new leukaemia model. In contrast to the *Mx-1-Cre* inducible system, in which *Pten* is deleted in nearly 100% of adult HSCs<sup>5,6</sup>, *Pten* in this new model is conditionally deleted in 40% of fetal liver HSCs and their differentiated progeny (W.G. and H.W., unpublished observations) by the expression of the *VE-cadherin-Cre* (*VEC-Cre*) transgene<sup>7</sup>. *VEC-Cre*-mediated PTEN loss leads to a progressive development of MPD in the chronic phase followed by blast crisis. As early as one month after birth, mutant mice developed a myeloid shift with increased neutrophil counts (Fig. 1b, P30, and Supplementary Fig. 1, CP). One to two months later, mutant mice showed a marked increase in circulating neutrophils and white blood cells (Fig. 1b, P60–P90) and leukaemic blast invasion into haematopoietic and non-haematopoietic organs

(Supplementary Fig. 1, BC). All mutant mice ( $n = 266$ ) died with hepatomegaly and splenomegaly, and 70% of them suffered from enlarged thymus and lymph nodes (Fig. 1a and data not shown). By adopting CD45/side-scatter (CD45/SSC) fluorescence-activated cell sorting (FACS) analysis, a methodology used for the characterization of human leukaemic blasts<sup>8</sup>, we further identified two leukaemia subtypes (Fig. 1c): T-ALL (CD3<sup>+</sup>CD4<sup>+</sup>/CD8<sup>+</sup>) in 74% of the mice and acute myeloid leukaemia (AML)/T-ALL (Gr-1<sup>low</sup>Mac-1<sup>+</sup> and CD3<sup>+</sup>CD4<sup>+</sup>) in the remaining 26%. This new model therefore shares a similar phenotype with the *Mx-1*-inducible model<sup>5,6</sup> but progresses at a much slower pace (three to four months instead of three weeks) and develops a predominant T-ALL phenotype.



**Figure 1 | *VEC-Cre*-mediated *Pten* loss leads to MPD and leukaemogenesis.** **a**, Survival curve for mouse littermate pairs. Inset: representative spleens from mice at postnatal day 60 (P60). Mut, mutant; WT, wild type. **b**, Progressive alterations in PB: 30 littermate pairs (W, wild-type; M, mutant) per time point. Bars show the data range; points show averages. WBC, white blood cells. **c**, Identification of blast population by CD45/SSC plot in BM and spleen. The blasts were sorted for Giemsa–Wright staining, PCR genotyping and lineage analyses.  $\Delta$ ex5, exon-5-deleted *Pten* allele.

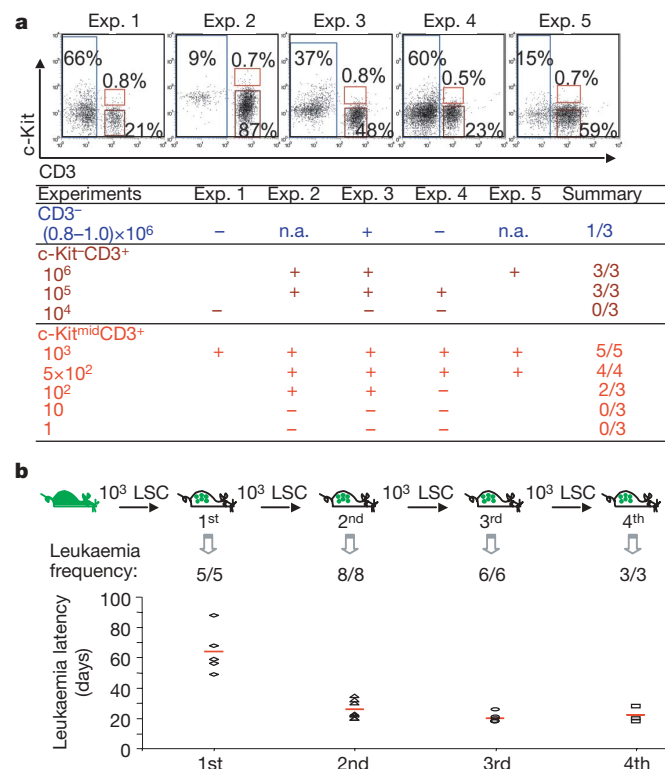
<sup>1</sup>Department of Molecular and Medical Pharmacology, <sup>2</sup>Department of Pediatric Hematology/Oncology, <sup>3</sup>Department of Molecular, and <sup>4</sup>Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Cellular and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095, USA. <sup>5</sup>Department of Medicine, Medical Oncology Division, University of Colorado Cancer Center, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA. <sup>6</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

To understand the molecular and genetic mechanisms involved in LSC formation, we sought to identify the LSC population in our model. We first determined whether leukaemia development was transplantable and autonomous for *Pten*-null cells by transplanting  $2 \times 10^6$  cells isolated from the bone marrow (BM), spleen or thymus of chronic-phase mutant male mice with the *ROSA26-LacZ* reporter gene<sup>9</sup> into sublethally irradiated mice with severe combined immunodeficiency disease (SCID) (Supplementary Fig. 2a). Because both *LacZ* reporter activation and the *Pten* gene deletion are controlled by the same *Cre* transgene, *LacZ* expression, measurable by a  $\beta$ -galactosidase fluorescent substrate and FACS analysis (FACS-Gal), was used to mark *Pten*-null cells in this and subsequent experiments. Our results demonstrated that 90% of the recipient mice ( $n = 22$ ) developed leukaemia within three months after transplantation. Strikingly, more than 95% of leukaemic blasts in SCID recipients were *LacZ*<sup>+</sup> and *Pten*-null, and infiltrated into non-haematopoietic organs (Supplementary Fig. 2b–d), in a similar manner to the primary disease (Supplementary Fig. 1). This result suggests that the leukaemia is initiated by donor-derived *Pten*-null LSCs.

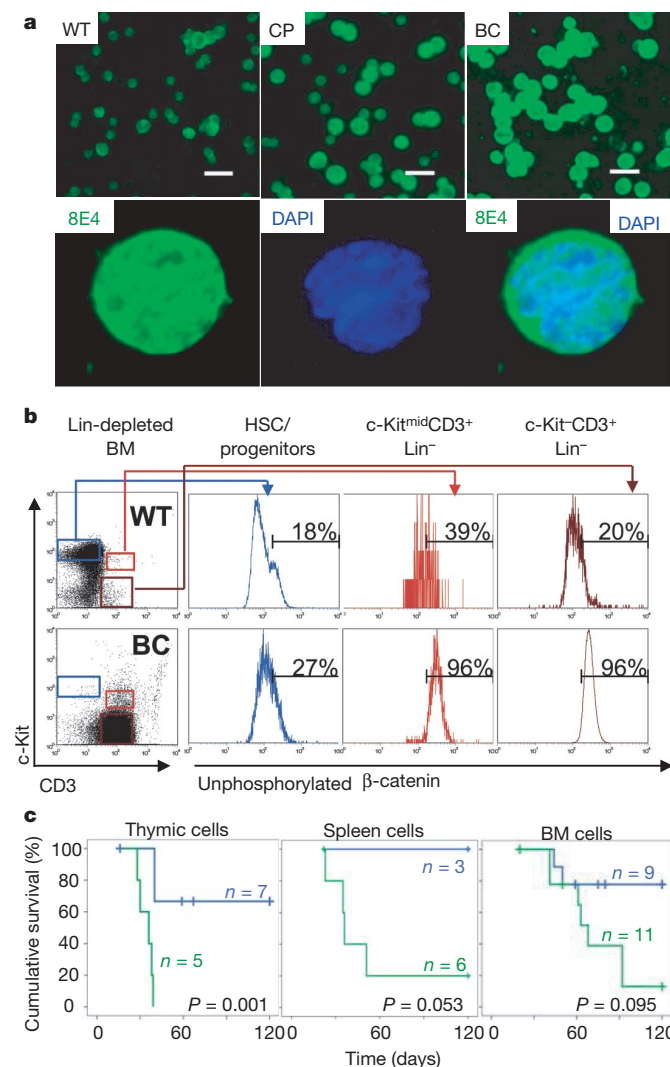
LSCs have been reported to arise from HSCs<sup>10,11</sup> or from myeloid and B-lineage progenitors<sup>12,13</sup>. Our initial screening of several stem-cell and progenitor markers identified a small c-Kit<sup>mid</sup> subgroup within the blast compartment (Supplementary Fig. 2e). When combined with CD3, a pan-T-cell marker, we further identified three major populations: CD3<sup>+</sup>, c-Kit<sup>mid</sup>CD3<sup>+</sup> and c-Kit<sup>mid</sup>CD3<sup>+</sup> (Supplementary Fig. 3). We then sorted these three subpopulations from an individual leukaemic mouse and performed transplantation experiments with various numbers of sorted cells (Fig. 2a and

Supplementary Fig. 3). When  $10^4$  c-Kit<sup>mid</sup>CD3<sup>+</sup> cells were transplanted into sublethally irradiated SCID mice, no leukaemia development was observed within 100 days, the longest time point we have followed for transplantation experiments. This is in sharp contrast with T-ALL development in those recipients transplanted with  $10^2$ – $10^3$  c-Kit<sup>mid</sup>CD3<sup>+</sup> cells. These results suggest that LSCs are enriched in a c-Kit<sup>mid</sup>CD3<sup>+</sup> subpopulation. Therefore, in contrast with the recent report<sup>14</sup>, leukaemia development in our *Pten*-null leukaemia model is driven by rare LSCs.

LSCs should be able to self-renew and initiate leukaemogenesis through serial transplantations. To test the self-renewal capability of the putative c-Kit<sup>mid</sup>CD3<sup>+</sup> LSCs, we transplanted 1,000 c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>−</sup> cells sorted from primary mice and then from each passage. The same immunophenotypic c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>−</sup> cells were able to mediate 100% leukaemia development in the quaternary



**Figure 2 | Self-renewing LSCs are enriched in the c-Kit<sup>mid</sup>CD3<sup>+</sup> compartment.** **a**, LSC identification. The experimental design is illustrated in Supplementary Fig. 3. Top: illustration of the three subpopulations that were sorted from BM of five independent leukaemic donor mice (cell fractions are denoted on each FACS plot). Bottom: summary of independent transplantation experiments with sorted and serially diluted cells. +, leukaemia development; –, leukaemia-free for more than 100 days; n.a., viable cells after sorting were not enough for transplantation. **b**, LSCs are self-renewing and lead to accelerated leukaemogenesis during serial transplantations. Top: illustration of the experimental procedure. Bottom: summary of the results. Red lines in the lower chart represent the means of leukaemia latencies.



**Figure 3 |  $\beta$ -Catenin activation in LSCs and its role in leukaemogenesis.** **a**, Accumulation of unphosphorylated  $\beta$ -catenin in blasts. Cytospin slides with thymic cells were stained with the monoclonal antibody 8E4, which recognizes unphosphorylated  $\beta$ -catenin. Top: representative fluorescent images. CP, chronic phase; BC, blast crisis. Scale bars, 25  $\mu$ m. Bottom: representative confocal images. DAPI, 4',6-diamidino-2-phenylindole. Original magnification  $\times 100$ . **b**, Marked increase in unphosphorylated  $\beta$ -catenin levels in the LSC and blast populations. BM cells were pooled from two blast-crisis or WT littermates and were lineage (Lin)-depleted before FACS analysis. **c**, Decreased and delayed leukaemogenesis after ablation of one allele of *Ctnnb1*. Kaplan–Meier survival curves with Logrank statistical analysis (denoted on the curves) summarize leukaemia development in transplantation experiments. Blue line, *Pten*<sup>loxP/loxP</sup>; *Ctnnb1*<sup>loxP/+</sup>; *VEC-Cre*<sup>+</sup>; green line, *Pten*<sup>loxP/loxP</sup>; *VEC-Cre*<sup>+</sup>.



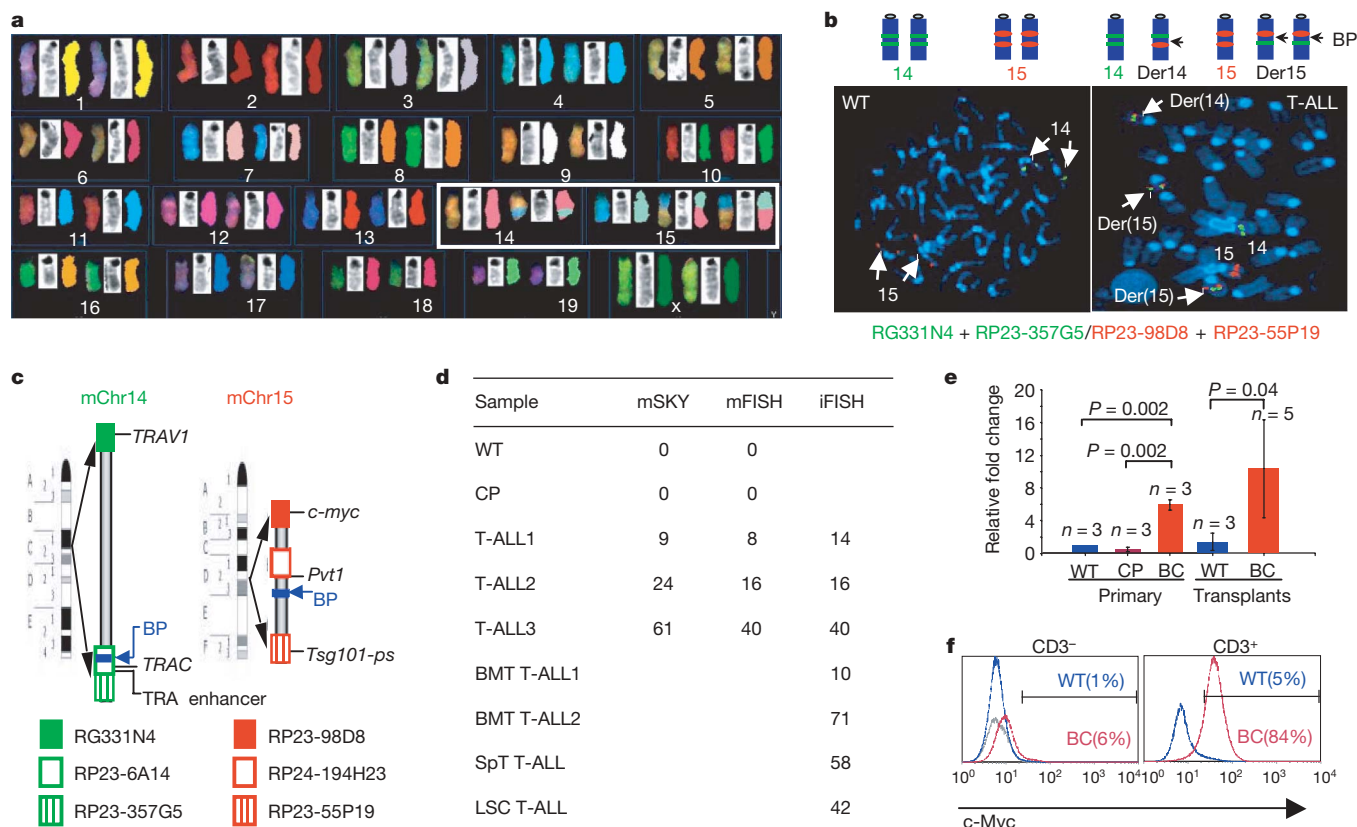
transplantation (Fig. 2b and Supplementary Fig. 4c). Although the same numbers of c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>-</sup> cells were serially transplanted, disease latencies were significantly shortened from 64 days to 20 days (Fig. 2b). Similar results were obtained when LacZ<sup>+</sup>-c-Kit<sup>mid</sup>CD3<sup>+</sup>-derived BM cells were serially transplanted (Supplementary Fig. 4a, b). Thus, similarly to a previous study of human LSCs<sup>10</sup>, LSCs in our model are self-renewable and give rise to a more accelerated disease through serial transplantations.

Because *Pten* deletion in HSCs results in a self-renewal defect and stem cell exhaustion<sup>5,6</sup> (W.G., J.L. and H.W., unpublished observations), understanding the molecular mechanism involved in the self-renewal acquisition of the c-Kit<sup>mid</sup>CD3<sup>+</sup> LSCs became the central focus of our study. The Wnt/ $\beta$ -catenin signalling pathway is known to be involved in the regulation of HSC self-renewal<sup>15</sup>, and its activation is required for the *in vitro* replating activity of the LSCs from myeloid blast crisis of human chronic myeloid leukaemia<sup>13</sup>. To assess  $\beta$ -catenin activation in our model, we employed a monoclonal antibody (8E4) that specifically recognizes the unphosphorylated and thus activated form of  $\beta$ -catenin. When making a comparison with the same immunophenotypic populations from wild-type (WT) mice, we detected a marked increase in unphosphorylated  $\beta$ -catenin in more than 90% of c-Kit<sup>mid</sup>CD3<sup>+</sup> LSCs and c-Kit<sup>+</sup>CD3<sup>+</sup> blasts (Fig. 3a, b, Supplementary Figs 5 and 6c, and data not shown) and modest increases in the c-Kit<sup>high</sup>Lin<sup>-</sup> HSC/progenitor population in the blast-crisis mice (Fig. 3b) and the CD3<sup>-</sup>CD4<sup>-</sup>CD8<sup>-</sup> T progenitors in the chronic-phase mice (Supplementary Fig. 6b). This result

suggests that activation of  $\beta$ -catenin may be associated with LSC formation and T-ALL development.

To improve our understanding of the role of  $\beta$ -catenin in LSC formation, we took a genetic approach by simultaneously deleting the genes encoding PTEN and  $\beta$ -catenin (*Pten* and *Ctnnb1*, respectively) in HSCs. The resulting *Pten*<sup>loxP/loxP</sup>; *Ctnnb1*<sup>loxP/loxP</sup>; *VEC-Cre*<sup>+</sup> mice are embryonic lethal. *Pten*<sup>loxP/loxP</sup>; *Ctnnb1*<sup>loxP/+</sup>; *VEC-Cre*<sup>+</sup> mice developed MPD with 100% penetrance at about 30 postnatal days (Supplementary Fig. 7a, b) but died with vascular complication before 68 postnatal days (data not shown). This prompted us to use the transplantation assay to compare the leukaemogenic potential between *Pten*<sup>loxP/loxP</sup>; *Ctnnb1*<sup>loxP/+</sup>; *VEC-Cre*<sup>+</sup> (blue line; three independent experiments) and *Pten*<sup>loxP/loxP</sup>; *VEC-Cre*<sup>+</sup> (green line; four independent experiments) mice. A gradient of leukaemogenic potential was observed in cells derived from different mutant haematopoietic organs (thymus > spleen > BM), which was consistent with the relative enrichment of LSCs in these organs. The occurrence of acute leukaemia and death were markedly decreased and delayed when one allele of *Ctnnb1* was deleted from donor cells (Fig. 3c and Supplementary Fig. 7c). These results suggest that dysregulated  $\beta$ -catenin activity may contribute to LSC formation and leukaemogenesis in our model.

Chromosomal translocations are frequently associated with human leukaemia and lymphoma. To investigate chromosomal abnormalities in our model, we employed cytogenetic approaches, including spectral karyotyping (SKY) and fluorescence *in situ*



**Figure 4 | The recurring translocation T(14;15) involves the *TcrA/TcrD* cluster and the *c-myc* gene and results in aberrant overexpression of *c-myc* in LSCs and T-ALL blasts.** **a**, Detection of T(14;15) by SKY analysis in metaphases prepared directly from BM of colcemid-treated WT or mutant mice. **b**, Breakpoint (BP) identification by two-colour FISH analysis. The results with green-labelled and red-labelled BAC probes are illustrated in the lower images and explained in the upper diagram. **c**, A schematic mapping of the T(14;15) translocation. BP, breakpoint. mChr, mouse chromosome. **d**, Recurring T(14;15) in primary T-ALL mutants and transplants with chronic-phase BM (BMT), splenic (SpT) or LSC cells was analysed by

metaphase SKY (mSKY), metaphase FISH (mFISH) and interphase FISH (iFISH). Values shown are percentages. **e**, *c-myc* messenger RNA overexpression (means  $\pm$  s.d.) in both primary blast-crisis (BC) mice and leukaemia transplants was detected by RT-PCR and analysed by Student's *t*-test. CP, chronic phase; C, control; WT littermates (for primary mice) or an unrelated SCID mouse (for WT and leukaemic transplants). **f**, Overexpression of *c-myc* only in CD3<sup>+</sup> splenic cells of T-ALL ( $n = 3$ ). The percentage with *c-myc* overexpression in the CD3<sup>-</sup> or CD3<sup>+</sup> compartment is denoted in parentheses. Grey, control (a mixture of WT and T-ALL cells); blue, WT; red, T-ALL.

hybridization (FISH) analyses. SKY analysis revealed no recurrent structural abnormalities in BM cells from chronic-phase mutant or WT mice (Fig. 4d). In contrast, all three blast-crisis samples we initially analysed showed the same T(14;15) chromosomal translocation in 9–61% of analysed metaphases (Fig. 4a, boxed chromosomes, and Fig. 4d), which was reminiscent of leukaemic clonal expansion in humans. Of the cells carrying T(14;15), 53% contained two or even three copies of Der(15)T(14;15) (Fig. 4a, b), suggesting a strong selective pressure in cells with this translocation and implying a possible role of translocation-associated genes in leukaemic transformation.

Because breakpoints in chromosomes 14 and 15 seemed to be located at 14qC2–C3 and 15qD3 (Fig. 4a), we searched for translocation targets in the public domain and identified the T-cell antigen receptor (TCR)- $\alpha/\delta$  (*Tcr $\alpha$ /Tcr $\delta$* ) cluster on chromosome 14qC2 and *c-myc* on chromosome 15qD3 as potential candidates. A similar translocation of t(8;14) is known to be associated with a subset of human T-ALL<sup>16</sup>. Using two-colour FISH analysis with validated bacterial artificial chromosome (BAC) clones (Supplementary Fig. 8a), we confirmed the translocation between *Tcr $\alpha$ /Tcr $\delta$*  and *c-myc* and narrowed down the breakpoints to two minimal regions (Fig. 4b, c, and Supplementary Fig. 8b): a roughly 209,000-base-pair (209-kb) fragment containing the *TRAC* gene and the TRA enhancer and 680 kb between the genes *Pvt1* and *Tsg101-ps*. The same translocation was found in four additional leukaemic recipients transplanted with primary chronic mutant cells or sorted LSCs (Fig. 4d and Supplementary Fig. 8c), suggesting that T-ALL can be recapitulated at the genetic level and that the genomic abnormality is intrinsic to LSCs.

To determine whether *c-myc* expression was indeed altered by *Tcr $\alpha$*  regulatory machinery, as reported for human T-ALL<sup>17</sup>, we quantified *c-myc* expression in BM and thymic cells with the use of quantitative RT–PCR and FACS analysis. Strikingly, *c-myc* expression was not changed in the chronic phase but was markedly increased in BM or thymic cells isolated from primary or transplanted blast-crisis mice (Fig. 4e). Furthermore, *c-myc* overexpression was detected only in LSCs and CD3<sup>+</sup> blast cells but not in HSCs or CD3<sup>–</sup> cells (Fig. 4f and Supplementary Fig. 9a, b). These results rule out a predominant role of the deregulated PTEN–PI(3)K–AKT pathway in *c-myc* overexpression and highlight the importance of overexpressed *C-myc* in the formation of CD3<sup>+</sup> LSCs and T-ALL development in our model.

Our study provides strong evidence that the molecular and genetic events involved in ‘multiple-hit’ leukaemogenesis are likely to take place at the levels of HSCs and LSCs<sup>18</sup>. In this model, *Pten* inactivation in HSCs serves as the first hit to activate the PI(3)K–AKT pathway, conferring survival and proliferative advantages, and to promote genomic instability<sup>19</sup>, leading to additional alterations. Among them, activation of  $\beta$ -catenin may contribute to the acquisition of self-renewal capacity of LSCs, while the T(14;15) chromosomal translocation results in T-lineage-specific overexpression of *c-myc*, which may promote *Pten*-null LSC self-renewal and lead to T-ALL development. The sequential order of  $\beta$ -catenin activation and *c-myc* overexpression is currently unknown but it does not seem that *c-Myc* function is mediated by  $\beta$ -catenin signalling (Supplementary Fig. 9c, d).

Dysregulated NOTCH1 signalling is involved in about 56% of human T-ALL<sup>3</sup>. However, in our leukaemia samples we detected no mutations in Notch1 heterodimerization (HD) domain and PEST domain, two mutation hot-spots associated with human T-ALL (Supplementary Fig. 10a). Furthermore, neither a consistent decrease in *Fbxw7* expression, a negative regulator of Notch signalling that is frequently altered in human T-ALL<sup>2</sup>, nor an increase in the expression of the NOTCH1 target gene *Hes1* was detected in this model (Supplementary Fig. 10b). Given recent reports that the activated Notch1 signalling directly induces *c-myc* overexpression<sup>20–22</sup> and modulates *Pten* expression<sup>23</sup>, the T(14;15)-mediated *c-myc* overexpression and *Pten* deletion together may substitute functionally for *Notch1* mutations, leading to T-ALL development in our model.

## METHODS SUMMARY

**Mice.** Mice were maintained in the animal facility of University of California Los Angeles in accordance with federal and institutional guidelines and were used between 1 and 5 months of age unless otherwise noted. *Pten*<sup>loxP/loxP</sup> mice have been deposited to the Jackson Laboratory.

**Transplantation assay.** Cells were directly collected from or sorted from BM, spleen or thymus, balanced with SCID carrier cells in some cases and transplanted into sublethally irradiated SCID recipients. Leukaemia was determined by CD45/SSC FACS, FACS–Gal and histological analyses.

**Flow cytometric (FACS) analysis.** Haematopoietic cells were obtained by flushing femurs or smashing the spleen and thymus, and were then stained with antibodies. These cells were either analysed or sorted by flow cytometry. Lineage depletion was performed with microbeads (Miltenyi Biotec). FACS–Gal analysis was performed by fluorescein di- $\beta$ -D-galactopyranoside loading and FACS analysis. To analyse activated  $\beta$ -catenin and *c-Myc*, cells were stained with surface markers, fixed, permeabilized and stained with corresponding antibodies for FACS analysis.

**Profiling and histology analysis of peripheral blood.** Peripheral blood (PB) was collected, analysed for PB profile, spread for blood smears and stained with Giemsa stain or Giemsa–Wright (Fisher). Tissues dissected from killed mice were fixed, embedded in paraffin and sectioned for haematoxylin/eosin staining. Cytospin slides with single suspension cells were prepared for immunocytochemistry with antibodies.

**SKY and FISH analysis.** Metaphases were prepared directly from BM cells in colcemid-treated mice for SKY assay and FISH analysis. Six murine DNA BAC clones were obtained, prepared, and validated for FISH analysis.

**RT–PCR.** Total RNA was extracted and reverse-transcribed into complementary DNAs. These cDNAs were diluted for real-time RT–PCR with primers for the genes of interest and reference. *C<sub>t</sub>* was obtained for a statistical calculation of relative fold changes.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 October 2007; accepted 25 March 2008.

Published online 7 May 2008.

- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
- Maser, R. S. *et al.* Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447**, 966–971 (2007).
- Weng, A. P. *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
- Fukuda, R. *et al.* Alteration of phosphatidylinositol 3-kinase cascade in the multilobulated nuclear formation of adult T cell leukemia/lymphoma (ATLL). *Proc. Natl Acad. Sci. USA* **102**, 15213–15218 (2005).
- Yilmaz, O. H. *et al.* *Pten* dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
- Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441**, 518–522 (2006).
- Alva, J. A. *et al.* VE-Cadherin-Cre-recombinase transgenic mouse: a tool for lineage analysis and gene deletion in endothelial cells. *Dev. Dyn.* **235**, 759–767 (2006).
- Borowitz, M. J., Guenther, K. L., Shultz, K. E. & Stelzer, G. T. Immunophenotyping of acute leukemia by flow cytometric analysis. Use of CD45 and right-angle light scatter to gate on leukemic blasts in three-color analysis. *Am. J. Clin. Pathol.* **100**, 534–540 (1993).
- Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
- Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
- Pasgue, E., Wagner, E. F. & Weissman Jun, I. L. B deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. *Cell* **119**, 431–443 (2004).
- Castor, A. *et al.* Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nature Med.* **11**, 630–637 (2005).
- Jamieson, C. H. *et al.* Granulocyte–macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N. Engl. J. Med.* **351**, 657–667 (2004).
- Kelly, P. N. *et al.* Tumor growth need not be driven by rare cancer stem cells. *Science* **317**, 337 (2007).
- Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).
- Erikson, J. *et al.* Deregulation of *c-myc* by translocation of the alpha-locus of the T-cell receptor in T-cell leukemias. *Science* **232**, 884–886 (1986).
- Boxer, L. M. & Dang, C. V. Translocations involving *c-myc* and *c-myc* function. *Oncogene* **20**, 5595–5610 (2001).
- Weissman, I. Stem cell research: paths to cancer therapies and regenerative medicine. *J. Am. Med. Assoc.* **294**, 1359–1366 (2005).

19. Shen, W. H. *et al.* Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell* **128**, 157–170 (2007).
20. Palomero, T. *et al.* NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).
21. Sharma, V. M. *et al.* Notch1 contributes to mouse T-cell leukemia by directly inducing the expression of *c-myc*. *Mol. Cell. Biol.* **26**, 8022–8031 (2006).
22. Weng, A. P. *et al.* c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev.* **20**, 2096–2109 (2006).
23. Palomero, T. *et al.* Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nature Med.* **13**, 1203–1210 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank O. Witte, K. Dorshkind, J. Said, R. Gatti, K. Sakamoto, S. Schubert, R. Hill and B. Valamehr for helpful comments; D. Cheng from O. Witte's laboratory for cell sorting; J. Gao for retro-orbital bleeding and genotyping; and the Department of Pathology's Tissue Procurement Core Laboratory (partly supported by National Cancer Institute grant CA16042) for tissue procurement. W.G. is supported by a California Institute of Regenerative

Medicine (CIRM) training fellowship. J.E.Y. was supported by the University of California Los Angeles Amgen Scholar Program. J.Y.C. is supported by the Beckman Undergraduate Research Program.

**Author Contributions** W.G. and H.W. designed the experiments, and W.G. performed a majority of the work involved in this study. J.L.L. performed immunohistochemistry on tissue sections and cytospin slides (Fig. 3a), and participated in FACS-Gal analysis, PB profile analysis and histological analysis. C.-J.C. performed RT-PCR and western blotting for *c-myc* expression (Fig. 4e and data not shown). X.L. maintained mouse colonies and was responsible for intravenous injection; S.M. performed western blotting and TOPflash reporter assay for  $\beta$ -catenin activation by c-Myc (Supplementary Fig. 9c, d). J.E.Y. performed *Notch1* mutation screening and RT-PCR analysis for *Fbxw7* and *Hes1* expression (Supplementary Fig. 10). J.Y.C. performed *Pten* and *Ctnnb1* genotyping and FACS analysis; Y.X. and M.V.-G. performed SKY and FISH analysis (Fig. 4a, b, d, and Supplementary Fig. 8). M.L.I.-A. provided *VE-Cadherin-Cre<sup>+</sup>* mice. W.G., J.L.L. and H.W. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.W. ([hwu@mednet.ucla.edu](mailto:hwu@mednet.ucla.edu)).



## METHODS

**Mice.** *VEC-Cre<sup>+</sup>* transgenic mice (129/C57 mixed background)<sup>7</sup> were carefully analysed and confirmed without any aberrant phenotypes before being crossed to *Pten<sup>loxP/loxP</sup>* mice (129/Balb/c mixed background)<sup>24</sup> to obtain *Pten<sup>loxP/+</sup>;VEC-Cre<sup>+</sup>* mice, which were backcrossed to *Pten<sup>loxP/loxP</sup>* mice to obtain *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>* mutant mice for this study. We also crossed *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>* mice with *ROSA26-LacZ<sup>+</sup>* reporter transgenic mice<sup>9</sup>, to obtain mice of *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>;LacZ<sup>+</sup>*, *Pten<sup>+/+</sup>;VEC-Cre<sup>+</sup>;LacZ<sup>+</sup>* and *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>;LacZ<sup>+</sup>* for transplantation and FACS-Gal analysis. *Ctnnb1<sup>loxP/loxP</sup>* mice<sup>25</sup> were crossed with *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>* mice to generate *Pten<sup>loxP/loxP</sup>;Ctnnb1<sup>loxP/+</sup>;VEC-Cre<sup>+</sup>* mice. Mouse genotypes were verified by PCR analysis with the primer sets for *Pten*, *Ctnnb1*, *Cre* and *LacZ* (Supplementary Fig. 11).

**Transplantation assay.** As illustrated in Supplementary Fig. 2a, cells harvested from BM, spleen or thymus of P30–P60 *Pten<sup>+/+</sup>;VEC-Cre<sup>+</sup>;LacZ<sup>+</sup>* WT, *Pten<sup>loxP/loxP</sup>;VEC-Cre<sup>+</sup>;LacZ<sup>+</sup>* or *Pten<sup>loxP/loxP</sup>;Ctnnb1<sup>loxP/+</sup>;VEC-Cre<sup>+</sup>* chronic mutant mice were injected into tail veins of 2–4-month-old CBySmn.CB17-Prkdc<sup>scid</sup> SCID female mice (Taconic), which had been sublethally irradiated at a dose of 180–200 rad on a <sup>137</sup>Cs Mark I irradiator (J. L. Shepherd & Associates). Development of MPD and leukaemia was monitored by FACS-Gal analysis on PB every 2 weeks.

For LSC identification, each of the three populations, as shown in Fig. 2b, was sorted from BM of blast-crisis mutants, serially diluted and balanced with *Pten<sup>+/+</sup>* carrier cells from SCID mice before being transplanted into sublethally irradiated SCID recipients. The relative fractions of these three populations varied from mouse to mouse. Leukaemia was determined by CD45/SSC FACS, FACS-Gal and histological analyses.

In serial transplantation assays, when a transplant recipient with 900 or 1,000 candidate LSCs (c-Kit<sup>mid</sup>CD3<sup>+</sup>Lin<sup>−</sup>) developed leukaemia, the same number of LSCs sorted (Fig. 2c) or various numbers of BM cells harvested (Supplementary Fig. 4c) from this recipient mouse were balanced with *Pten<sup>+/+</sup>* SCID carrier cells in most cases (indicated in the figure legends) and transplanted into secondary SCID recipients. The same procedure was repeated up to the fourth passage.

**Flow cytometric (FACS) analysis.** The cells stained with antibodies were either analysed by flow cytometry on a BD FACSCalibur or FACScan flow cytometer (BD Biosciences) with modifications by Cytek or sorted on a BD FACS Vantage s.e. sorter (BD Biosciences). Cells were stained with fluorescein isothiocyanate (FITC)-conjugated, R-phycoerythrin (PE)-conjugated, allophycocyanin (APC)-conjugated or APC-Cy7-conjugated antibodies, including Ter119, Gr-1 (RB6-8C5), Mac-1 (M1/70), CD3 (145-2C11), CD4 (GK1.5), CD8 (53-6.7), B220 (RA3-6B2), CD19 (1D3), c-Kit (2B8), Sca-1 (E13-161.7) or CD45 (30-F11) (from BD Pharmingen or eBioscience). For partial lineage depletion, cells were stained with lineage markers (PE–Ter119 and B220), followed by anti-PE microbeads (Miltenyi Biotec) and sorted for Lin<sup>low/−</sup> population in accordance with the manufacturer's instruction. Fluorescein di(β-D-galactopyranoside) (FDG) was obtained from Invitrogen or Sigma and used for FACS-Gal analysis in accordance with the manufacturers' instructions.

Leukaemic blasts were analysed on CD45/SSC plots as described previously<sup>8</sup>. An abnormal blast population was detected in all mutant animals at blast crisis and constituted more than 20% of the total leukocytes (Fig. 1c), satisfying the French–American–British (FAB) criteria for human acute leukaemia. The blast cells were *Pten*-null, large, immature and morphologically distinct from normal lineage cells but similar to human leukaemic blasts (Fig. 1c). Leukaemia with blasts positive for CD3, CD4 and/or CD8 was considered T-ALL, whereas leukaemia with more than 3% of Gr-1<sup>low</sup>Mac-1<sup>+</sup> blasts was AML. Blasts with both CD3<sup>+</sup>CD4<sup>+</sup> and Gr-1<sup>low</sup>Mac-1<sup>+</sup> may be either a mixture of T-ALL and AML or T-ALL bearing myeloid markers<sup>26</sup>.

Intracellular staining with an anti-β-catenin (unphosphorylated) mouse monoclonal antibody (clone 8E4; AXXORA or Millipore/Upstate), anti-c-myc rabbit polyclonal antibody (Cell Signaling Technology) or control mouse or rabbit IgG, followed by staining with donkey anti-mouse-FITC or anti-rabbit-FITC antibody (Jackson ImmunoResearch Laboratories), was performed with the Fix & Perm Kit (Caltag) in accordance with the manufacturer's instructions.

**PB profiling and histology analysis.** For PB profile analysis, 200 μl of PB was collected by retro-orbital bleeding and analysed on a HemaVet HV950FS (Drew

Scientific). Blood smears with eye or tail PB were subjected to Giemsa (Sigma) or Giemsa–Wright (Fisher) staining, in accordance with the manufacturers' instructions.

For histological analysis, tissues were fixed with Z-Fix (Anatech) or 10% formalin (Fisher) for 12 h, embedded with paraffin and sectioned for haematoxylin and eosin staining at the Tissue Procurement and Histology Core Laboratory at UCLA. Cytospin slides prepared with single suspension cells on a Shandon Cytospin 2 (Thermo) were fixed in 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, blocked with Mouse-on-Mouse blocking solution (Vector Laboratories) and immunostained with anti-un-phosphorylated β-catenin antibody (clone 8E4), anti-mouse IgG-biotin, and streptavidin-FITC (Vector Laboratories). Images were taken with a Macrofire charge-coupled device camera (Optronics) under a BX60 microscope (Olympus).

**SKY and FISH analysis.** Metaphases of BM cells were prepared directly from colcemid-treated mice to avoid any artificial chromosomal abnormality introduced by *in vitro* cell culture. To prepare metaphases, 250 μl of 200 μg ml<sup>−1</sup> colcemid (Sigma) was injected intraperitoneally into mice. After 30 min, BM cells were flushed out with 30 ml of 0.06 M potassium chloride, lysed at 37 °C for 20 min and fixed with 20 ml of fixative (3:1 methanol/acetic acid) for 10 min at 22–25 °C. The SKY assay was performed in accordance with the manufacturer's instructions (Applied Spectral Imaging) and images were acquired using a SD300 Spectracube system (Applied Spectral Imaging) mounted in an Olympus BX60 microscope equipped with a custom-designed optical filter (SKY-1; Chroma Technology). The conversion of the emission spectra to the display colours was achieved by assigning blue, green and red colours to specific sections of the emission spectrum. For FISH analysis, murine DNA BAC clones RP23-6A14, RP23-357G5, RP23-98D8, RP24-194H23 and RP23-55P19 were obtained from the Children's Hospital Oakland Research Institute, and RG331N4 was obtained from Invitrogen. Their genomic sequences are available in the National Center for Biotechnology Information mouse genome resources (<http://www.ncbi.nlm.nih.gov/genome/guide/mouse/>). The BAC DNAs were prepared with a Qiagen Large Construction Kit (Qiagen) and labelled with SpectrumGreen-conjugated dUTPs (RG331N4, RP23-6A14 and RP23-357G5 on chromosome 14) or SpectrumRed-conjugated dUTPs (RP23-98D8, RP24-194H23 and RP23-55P19 on chromosome 15) using the Vysis Nick Translation Kit (Abbott Molecular), in accordance with the manufacturers' instructions. Single-colour and dual-colour FISH assays were performed as described previously<sup>27</sup>. FISH images were captured with a charge-coupled device camera under a Zeiss Axio Imager Z1 fluorescence microscope equipped with proper filters (Zeiss).

**Quantitative RT-PCR.** Total RNA was extracted with a Qiagen RNeasy Micro RNA Kit (Qiagen) from mouse BM, splenic cells and thymic cells, or FACS-sorted cells from BM, and reverse-transcribed into cDNA with Superscript III Reverse Transcriptase (Invitrogen), in accordance with the manufacturers' instructions. cDNA was diluted 1:20 and mixed with the SYBR Green I mix (Bio-Rad) to perform RT-PCR with primers for the target genes *c-myc*, *Fbxw7* or *Hes1* or the reference gene for β-actin (*Actb*; sequences are given in Supplementary Fig. 11) in an iCycler (Bio-Rad). Amplification of correct products with no genomic or non-specific noise was verified on agarose gels. Each reaction was repeated three or four times to ensure *C<sub>t</sub>* consistency (a *C<sub>t</sub>* variation of one cycle or less), and the mean *C<sub>t</sub>* was used for statistical calculation of relative fold changes. The relative fold change for each target mRNA was calculated as  $2^{\Delta C_{t, \text{target}}(\text{control} - \text{sample}) - \Delta C_{t, \text{reference}}(\text{control} - \text{sample})}$ . The control samples were obtained from either unrelated WT mice or WT littermates (see figure legends).

24. Lesche, R. *et al.* Cre/loxP-mediated inactivation of the murine Pten tumor suppressor gene. *Genesis* **32**, 148–149 (2002).
25. Brault, V. *et al.* Inactivation of the β-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* **128**, 1253–1264 (2001).
26. Putti, M. C. *et al.* Expression of myeloid markers lacks prognostic impact in children treated for acute lymphoblastic leukemia: Italian experience in AIEOP-ALL 88–91 studies. *Blood* **92**, 795–801 (1998).
27. Boomer, T. *et al.* Detection of E2A translocations in leukemias via fluorescence *in situ* hybridization. *Leukemia* **15**, 95–102 (2001).

## LETTERS

# Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes

Oliver H. Tam<sup>1\*</sup>, Alexei A. Aravin<sup>1\*</sup>, Paula Stein<sup>2</sup>, Angelique Girard<sup>1</sup>, Elizabeth P. Murchison<sup>1</sup>, Sihem Cheloufi<sup>1</sup>, Emily Hodges<sup>1</sup>, Martin Anger<sup>2†</sup>, Ravi Sachidanandam<sup>1</sup>, Richard M. Schultz<sup>2</sup> & Gregory J. Hannon<sup>1</sup>

Pseudogenes populate the mammalian genome as remnants of artefactual incorporation of coding messenger RNAs into transposon pathways<sup>1</sup>. Here we show that a subset of pseudogenes generates endogenous small interfering RNAs (endo-siRNAs) in mouse oocytes. These endo-siRNAs are often processed from double-stranded RNAs formed by hybridization of spliced transcripts from protein-coding genes to antisense transcripts from homologous pseudogenes. An inverted repeat pseudogene can also generate abundant small RNAs directly. A second class of endo-siRNAs may enforce repression of mobile genetic elements, acting together with Piwi-interacting RNAs. Loss of Dicer, a protein integral to small RNA production, increases expression of endo-siRNA targets, demonstrating their regulatory activity. Our findings indicate a function for pseudogenes in regulating gene expression by means of the RNA interference pathway and may, in part, explain the evolutionary pressure to conserve argonaute-mediated catalysis in mammals.

Small-RNA-directed gene silencing pathways have been adapted to accept numerous inputs and to act on many types of downstream targets. In few places is this more apparent than in animal germ lines where two classes of small RNAs—microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs)—with distinct biogenesis mechanisms and biological functions have been reported. Although miRNAs, as a group, are ubiquitously expressed, piRNAs have thus far been found only in germ cells and in a few gonadal somatic cells types<sup>2</sup>. piRNAs repress the activity of mobile genetic elements, forming a small RNA-based, innate immune system with both genetically encoded and adaptive components<sup>2–9</sup>.

In mice, a homozygous mutation in any single Piwi family member causes male sterility accompanied by gonadal hypotrophy<sup>5,10,11</sup>. In Mili and Miwi2 mutants, meiosis is not completed and germ cells are progressively lost<sup>5</sup>. This correlates with an activation of transposons, particularly the non-long terminal repeat (LTR) retrotransposon, L1 (refs 5, 12). DNA methylation of L1 elements is correspondingly lost. In contrast, females bearing homozygous mutations in individual Piwi genes are apparently normal and fertile<sup>5,10,11</sup>. Because female germ cells must also control transposons, we sought to characterize their small RNA profiles to determine whether a piRNA system, similar to that operating in spermatocytes, also exists in oocytes.

Approximately 6,000 fully grown oocytes, arrested in prophase of meiosis I, were collected. Small RNA fractions from 19–24 nucleotides (lower fraction) and 24–30 nucleotides (upper fraction) were gel purified and used to prepare small RNA libraries. These were deeply sequenced<sup>3,4</sup>. A total of 1,037,355 sequences was obtained that could be mapped to the mouse genome (753,981 from the lower

fraction and 283,374 from the upper fraction; Supplementary Table 1). In the lower fraction, 126,515 non-redundant sequences were obtained, falling into 24,271 non-overlapping clusters. In the upper fraction, 97,807 non-redundant sequences fell into 15,032 non-overlapping clusters.

An examination of the small RNAs in the upper fraction of the oocyte library revealed a piRNA population that resembled those found in early-stage spermatocytes<sup>3</sup> (Fig. 1a, right). Roughly 62% of small RNAs correspond to annotated repeats (Supplementary Table 2), with 3% matching genic sequences and 3% matching unannotated, intergenic sites. The function of the latter species remains unknown. Roughly 30% of the library corresponded to presumed breakdown products of abundant, non-coding RNAs, such as ribosomal RNAs, transfer RNAs and small nucleolar RNAs.

As expected, oocyte piRNAs arise from discrete genomic loci in a strand-asymmetric fashion (Supplementary Table 3). A number of these loci share structural similarities to *Drosophila* piRNA loci, which act as master controllers of mobile elements<sup>3</sup>. One example (Fig. 1b) spans ~120 kb of chromosome 10 and contains an abundance of long interspersed elements (LINEs) and LTR elements. These have an orientation bias that results in the generation of predominantly antisense piRNAs (Fig. 1b, piRNA, weighted; see also Supplementary Fig. 1).

piRNAs have been proposed to act with transcripts from active transposons in a feed-forward amplification loop that confers signature features on a piRNA population that is mounting an ongoing transposon defence<sup>2–4,6,7</sup>. Primary piRNA-directed cleavage of transposon mRNAs creates the 5' ends of secondary piRNAs<sup>4,6</sup>. This produces primary and secondary piRNA pairs that overlap by 10 nucleotides at their 5' ends. The 5' U bias of primary piRNAs thus leads to an enrichment of an A at position 10 of secondary piRNAs. These characteristics are prevalent in piRNA populations from mouse oocytes, particularly those that can be mapped to the L1 and intracisternal A particle (IAP) elements (Supplementary Fig. 2).

As expected, annotated miRNAs comprised the majority (69%) of 19–24-nucleotide RNAs (Fig. 1a, left; see also Supplementary Table 4). Among the highly abundant species are members of the let-7 family (let-7a/c/f), generally abundant miRNAs (miR-22, miR-16, miR-21, miR-26, miR-93 and miR-29a/b), and miRNAs abundant in ovary and placenta (miR-322, miR-503, miR-451). Finally, we detected miRNAs specifically expressed in male and female gonad (miR-103)<sup>13</sup>.

A substantial fraction of 19–24-nucleotide RNAs matched annotated transposons (Supplementary Table 2). Many that mapped uniquely to the genome could be traced to oocyte piRNA loci (Fig. 1b, siRNA, weighted). These species might represent piRNA

<sup>1</sup>Cold Spring Harbor Laboratory, Watson School of Biological Sciences and Howard Hughes Medical Institute, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

<sup>2</sup>Department of Biology, University of Pennsylvania, 433 South University Avenue, 205 Lynch Laboratories, Philadelphia, Pennsylvania 19104-6018, USA. †Present address: University of Oxford, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK.

\*These authors contributed equally to this work.

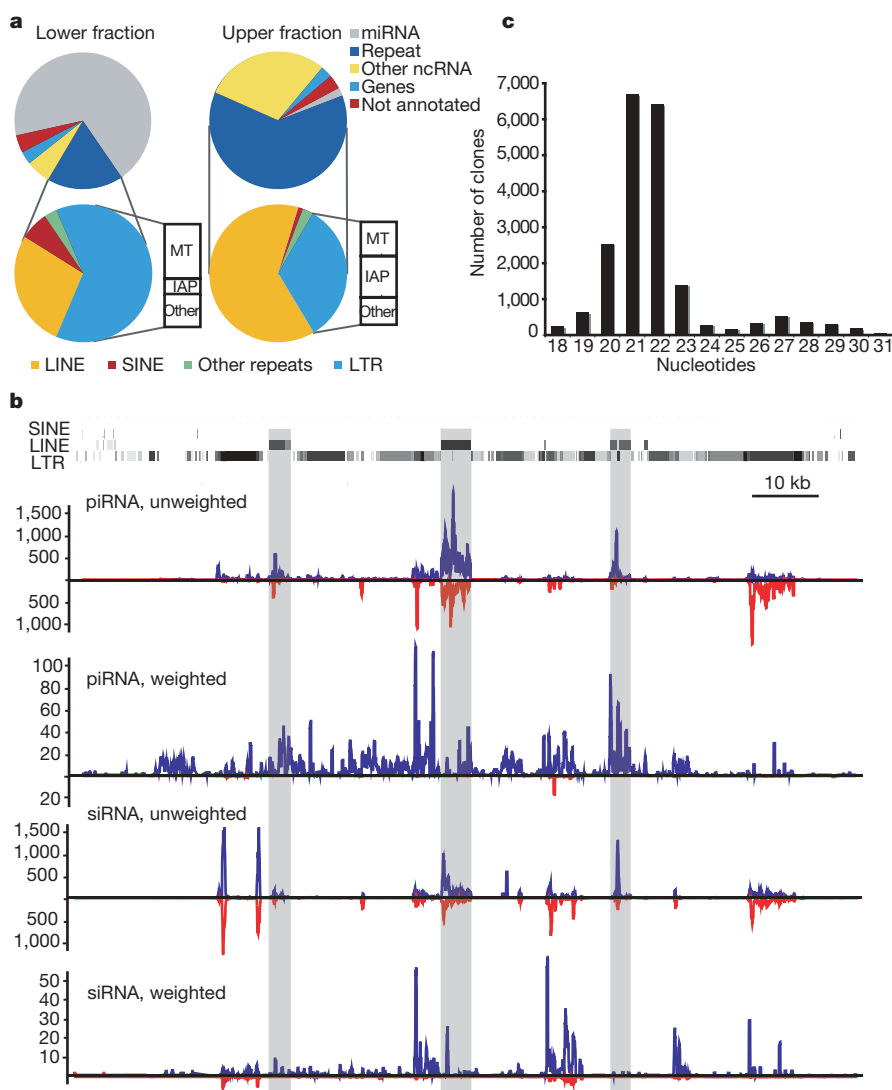
degradation products, or oocyte piRNA clusters might generate both siRNAs and piRNAs.

Therefore, we independently mapped piRNAs and candidate siRNAs to consensus L1 and IAP sequences (Supplementary Fig. 3). Each gave characteristic profiles. Moreover, piRNAs and candidate siRNAs show distinctly different nucleotide biases, with piRNAs displaying their characteristic enrichment for a 5' uridine residue and an A at position 10 (Supplementary Fig. 2). Candidate siRNAs lack a 10A bias and show enrichment for both A and U residues at their 5' ends (Supplementary Fig. 2). Finally, we gel purified 19–30-nucleotide RNAs from mouse oocytes as a single fraction and deeply sequenced this population. A length distribution of small RNAs that match the piRNA cluster shown in Fig. 1b yields two distinct peaks (Fig. 1c). siRNAs 21–22 nucleotides in length apparently predominate over the piRNA population, which averages ~27 nucleotides. We conclude that transposon-rich loci in oocytes give rise to both siRNAs and piRNAs. Although siRNAs are apparently more abundant, piRNA cloning frequencies could be reduced by the 2'-O-methyl modification that occurs on their 3' termini<sup>2</sup>. Our results

raise the possibility that piRNA and siRNA systems may act redundantly to repress transposons in mouse oocytes, perhaps explaining the lack of substantial phenotypic consequences of individual Piwi mutations in females<sup>5,10,11</sup>.

Although many transposons were targeted by both piRNAs and siRNAs, some relied more heavily on a particular pathway. For example, MTB and MTC, both LTR retrotransposons, matched almost exclusively to siRNAs. Moreover, the most prominent cluster that produces MTB/MTC small RNAs contains an inverted repeat with strong potential to produce a Dicer substrate. Notably, this transposon class showed increased expression in Dicer-null oocytes, consistent with its being regulated predominantly if not exclusively by the siRNA system<sup>14</sup>.

Small RNA libraries often contain genic sequences. In other tissues, these correspond exclusively to sense sequences that probably represent contaminating degradation products. However, in oocytes, numerous sense and antisense siRNAs corresponding to protein-coding genes could be identified (Supplementary Table 5). As mammals lack any identifiable RNA-dependent RNA



**Figure 1 | Both piRNA and siRNA systems control transposons in mouse oocytes.** **a**, Small RNA libraries from 19–24-nucleotide (lower fraction, left) and 24–30-nucleotide RNAs (upper fraction, right) were deeply sequenced. Reads were assigned an annotation as previously described<sup>12</sup>. The fraction of reads in each category is depicted. The repeat-annotated small RNAs were designated as LINE, SINE (short interspersed element), LTR and other, with the LTR category further divided between MT (mouse transcript), IAP and other. **b**, A representative piRNA locus on chromosome 10 is shown with the

content of LINE, SINE and LTR fragments. Shading indicates the degree of match to the consensus element. Frequency plots for piRNAs and siRNAs from this locus are shown below. 'Unweighted' plots each match to the cluster; 'weighted' normalizes each match, dividing by its genomic frequency<sup>4</sup>. Blue and red lines indicate small RNAs mapping to the upper and lower genomic strand, respectively. **c**, RNAs from 19–30 nucleotides were deeply sequenced. Reads matching the piRNA cluster shown in **b** were used to construct a frequency plot by length.



polymerase, this raised the question of how antisense siRNAs might be generated.

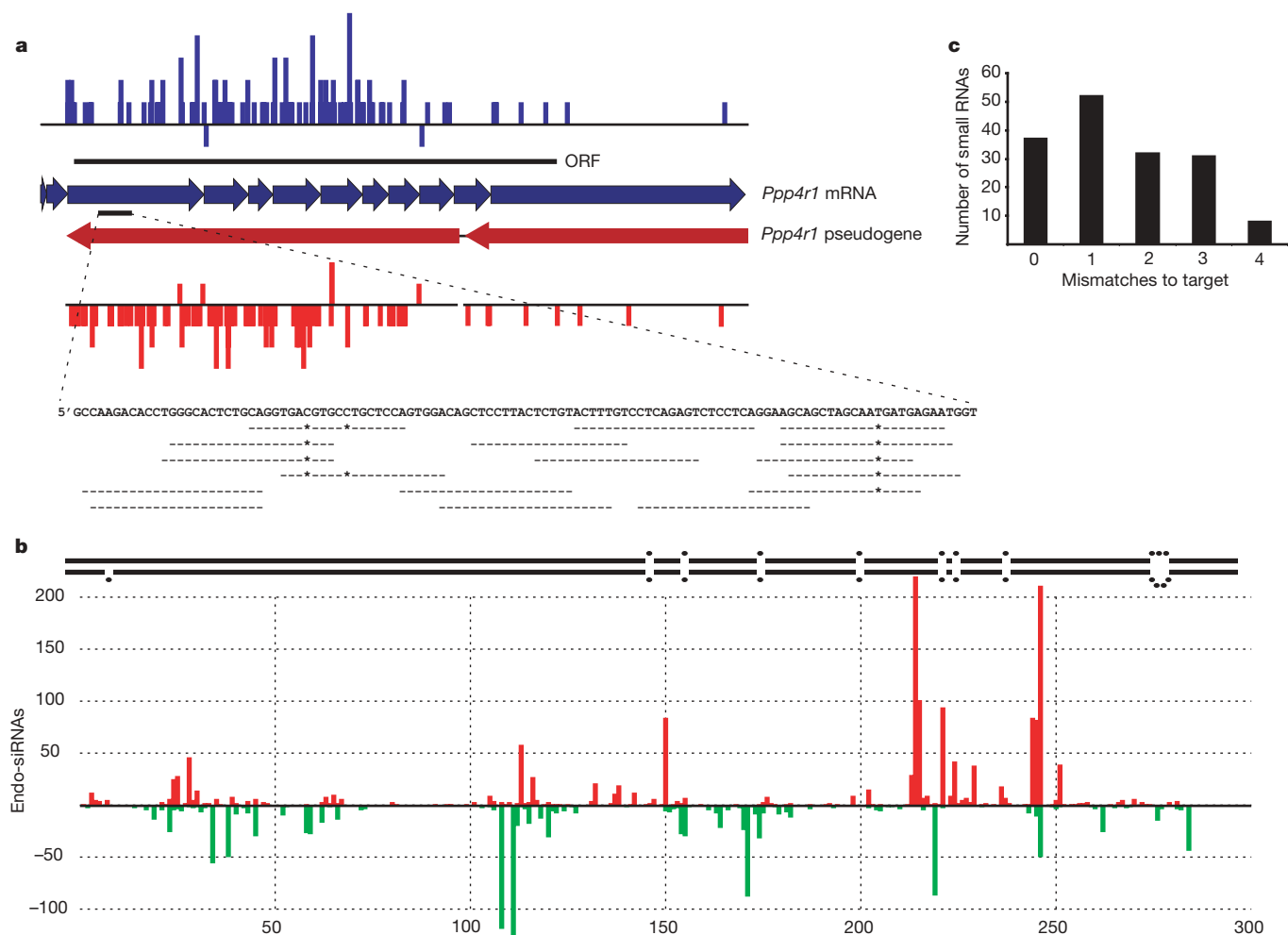
On the basis of polymorphisms, uniquely mapping sense siRNAs could often be assigned to the functional protein-coding copy of a gene, whereas antisense siRNAs mapped to a homologous pseudogene (Fig. 2a and Supplementary Table 5). Thus, oocyte endo-siRNAs might be processed from double-stranded (ds)RNAs that form by hybridization of transcripts derived from two unlinked loci. A similar process in which transcripts from active transposons hybridize to antisense transposon fragments transcribed from piRNA clusters could explain the genesis of transposon siRNAs.

siRNAs from gene–pseudogene pairs arise exclusively from regions of complementarity between the partners. Because many sense-oriented siRNAs cross exon–exon junctions, we propose that mature, spliced mRNAs from genes interact with antisense pseudogene transcripts to form Dicer substrates (Supplementary Fig. 4). In one case (Fig. 2b), both sense and antisense siRNAs to the GTPase-activating protein for Ran (Ran-GAP) were produced from a pseudogene locus containing a ~300-base pair (bp) inverted repeat with an intervening ~800-base loop. siRNAs were derived only from the potentially double-stranded portion of this locus.

In some cases, Dicer processing of dsRNA substrates proceeds in an apparently processive fashion from a discrete initiation site, producing ‘phased’ small RNAs with an ~21-nucleotide periodicity<sup>15</sup>. Transposon-derived and genic siRNAs showed this property only very weakly (Supplementary Fig. 2). Notably, piRNAs also show a similar, very weak phasing signal, although with a period of ~27 nucleotides rather than ~21 nucleotides.

Pseudogenes have often diverged substantially from their functional ancestors. Thus, we wished to examine the possibility that pseudogene-derived antisense siRNAs could regulate corresponding protein-coding genes. We mapped antisense siRNAs to potentially relevant regulatory targets. Many small RNAs aligned to their targets either with no mismatches or with mismatches lying outside regions essential for slicer cleavage<sup>16,17</sup> (Fig. 2a, c). Thus, antisense, pseudogene-derived siRNAs might be capable of regulating homologous protein-coding genes through a conventional RNA-interference mechanism.

To test the regulatory potential of pseudogene-derived siRNAs, we assessed the effects of Dicer loss on their putative targets. We have previously shown that deletion of *Dicer* in growing oocytes causes the production of non-functional gametes with defects in spindle organization and chromosome segregation<sup>14,18</sup>. We compared the



**Figure 2 | Gene–pseudogene interactions produce endogenous siRNAs.** **a**, Endo-siRNAs unambiguously mapped to the functional *Ppp4r1* mRNA are plotted in blue, above the mRNA (individual exons indicated as thick arrows). The extent of the open reading frame (ORF) is indicated. Endo-siRNAs from the *Ppp4r1* pseudogene are shown in red below the pseudogene. Arrows indicate two segments of *Ppp4r1* homology. siRNAs plotted above each line are sense oriented, with respect to the functional mRNA, and those below are antisense. Shown below is an enlargement of one section of the mRNA with individual antisense siRNAs aligned. A dash

indicates a match; an asterisk indicates a mismatch. The heights of the bars indicate the number of siRNAs starting at each position. **b**, Endogenous siRNAs homologous to Ran-GAP are shown below a schematic of the genomic inverted repeat structure from which they arise (the ~800-base loop is not depicted). Those shown above and below the x axis come from the upper or lower arm of the hairpin, respectively. No siRNAs were sequenced from the ~800-base loop that separates the two stem arms. **c**, siRNAs antisense to *Ppp4r1* counted and plotted by the number of mismatches to the functional mRNA.

expression of candidate endo-siRNA targets in wild-type and Dicer-null cells<sup>14</sup>. Many genes with abundant, pseudogene-derived siRNAs showed significant increases in expression following Dicer loss (Fig. 3a). We verified candidates derived from the array data by semi-quantitative polymerase chain reaction with reverse transcription (qRT-PCR, Fig. 3b).

Collectively, our data indicate that in mammalian oocytes, protein-coding mRNAs interact with pseudogene transcripts to form dsRNAs that are processed into endo-siRNAs. Examination of *Dicer* knockouts indicates a function for endo-siRNAs in gene regulation. At present, we cannot distinguish whether these small RNAs direct target cleavage or whether the act of siRNA production *per se*, which consumes the coding mRNA, is sufficient for repression. However, the specific case of HDAC1 may point to a RNA-induced silencing complex (RISC)-based mechanism. Few uniquely mapping

siRNAs are generated from the *Hdac1* gene itself, suggesting that it is not used prominently as a Dicer substrate. Instead, most uniquely mapping sense and antisense siRNAs can be assigned to a series of *Hdac1* pseudogenes. On the basis of its increased expression in Dicer-null oocytes, we propose that pseudogene-derived, antisense siRNAs direct RISC to cleave *Hdac1* mRNAs.

The catalytic potential of at least one argonaute protein has been conserved through mammalian evolution from platypus to humans<sup>19,20</sup>. However, mammalian miRNAs, with one known exception, act through translational mechanisms without the need for mRNA cleavage<sup>21</sup>. The discovery of endogenous siRNAs in mammalian oocytes not only expands the realm of mammalian small RNA classes but also provides one possible explanation for the evolutionary pressure to conserve argonaute enzymatic activity.

Pseudogenes have long been considered to be non-functional artefacts of transposition pathways acting on protein-coding mRNAs. In a few cases, regulatory roles have been posited for pseudogenes, largely through antisense mechanisms<sup>22–24</sup>. Our findings, and those of the accompanying paper<sup>25</sup>, provide a role for a subset of mammalian pseudogenes in the production of functional siRNAs. The production of dsRNAs by interaction between sense and antisense transcripts from distinct loci has not been observed in other tissues and may require the unique environment of oocytes, which substantially lack a protein kinase R response (a dsRNA-induced general translational repression pathway) and are geared for mRNA stabilization and storage<sup>26–28</sup>. The fact that many targets of this pathway are related to microtubule dynamics (including microtubule-based processes,  $P = 0$ ; kinesin complex,  $P = 0$ ; motor activity,  $P < 1 \times 10^{-254}$ ; spindle,  $P < 8 \times 10^{-239}$ ; and microtubule-associated complex,  $P < 3 \times 10^{-60}$ ; Supplementary Fig. 5) suggests that the regulatory circuits that we describe may have important biological roles, as the consequences of *Dicer* loss in growing oocytes is disruption of proper spindle formation and defects in chromosome segregation<sup>14,18</sup>.

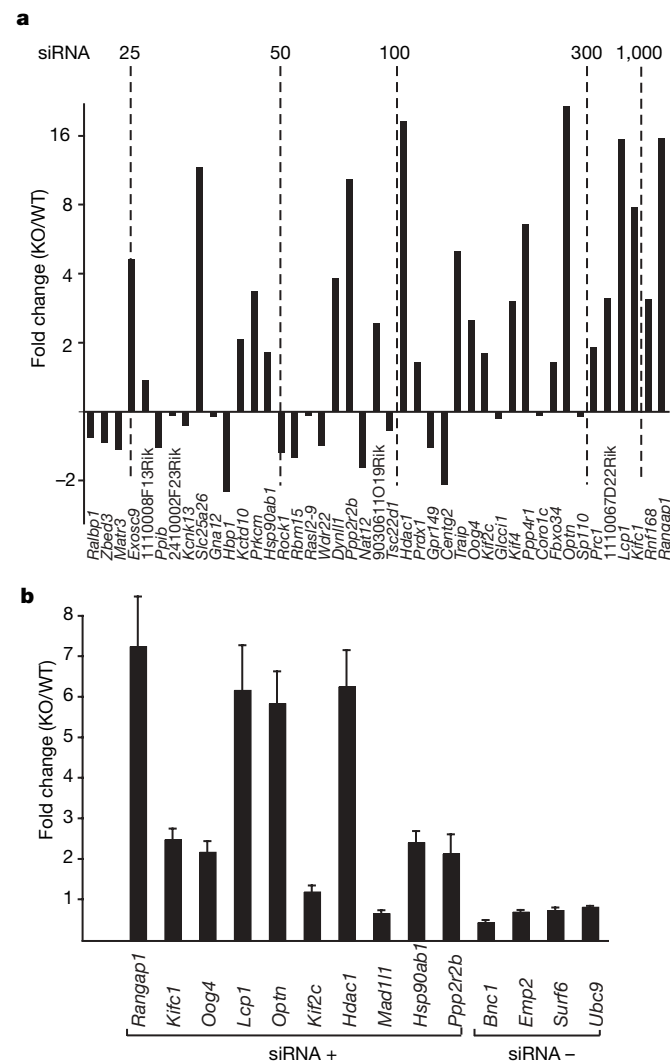
## METHODS SUMMARY

Mouse oocytes were collected from primed mice and used to prepare small RNA fractions. These were cloned and deeply sequenced as previously described<sup>24</sup>. Bioinformatic analysis of the sequences was performed as described in the Methods. For semi-quantitative RT-PCR, RNA was extracted from fully grown oocytes from *Dicer*<sup>flax/flax</sup> and *Dicer*<sup>flax/flax</sup> *Zp3-cre* mice. Quantitative PCR was performed using TaqMan probes.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 1 November 2007; accepted 7 March 2008.

Published online 10 April 2008.



**Figure 3 | Endo-siRNAs have a role in gene regulation.** **a**, We had previously compared expression levels in wild-type and Dicer-null oocytes by microarray<sup>14</sup>. Genes with a large number of siRNAs were examined. For those with significant changes in expression ( $P < 0.1$ ), the fold change in Dicer-null (KO) versus wild type (WT) was plotted. The graph was arranged according to the number of antisense siRNAs per gene in our data set, increasing from left to right (with benchmarks shown). The identity of the gene represented by each bar is given below. **b**, Fourteen genes (with and without siRNAs in our data set, indicated by siRNA+ and siRNA-, respectively), comprising a set in which some showed significant changes and some did not, were tested for changes in mRNA levels in Dicer-null versus wild-type oocytes by quantitative PCR. Gene names are indicated below each bar; error bars represent the mean  $\pm$  s.e.m. (three replicates).

1. D'Errico, I., Gadaleta, G. & Saccone, C. Pseudogenes in metazoa: origin and features. *Brief. Funct. Genomics Proteomics* **3**, 157–167 (2004).
2. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
3. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).
4. Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
5. Carmell, M. A. et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).
6. Gunawardane, L. S. et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
7. Houwing, S. et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**, 69–82 (2007).
8. Saito, K. et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20**, 2214–2222 (2006).
9. Vagin, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
10. Kuramochi-Miyagawa, S. et al. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).
11. Deng, W. & Lin, H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2**, 819–830 (2002).
12. Aravin, A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).

13. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
14. Murchison, E. P. *et al.* Critical roles for Dicer in the female germline. *Genes Dev.* **21**, 682–693 (2007).
15. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
16. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
17. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg<sup>2+</sup>-dependent endonuclease. *Curr. Biol.* **14**, 787–791 (2004).
18. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21**, 644–648 (2007).
19. Joshua-Tor, L. The Argonautes. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 67–72 (2006).
20. Murchison, E. P. *et al.* Conservation of small RNA pathways in platypus. *Genome Res.* (in the press).
21. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
22. Korneev, S. A., Park, J. H. & O'Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* **19**, 7711–7720 (1999).
23. Weil, D., Power, M. A., Webb, G. C. & Li, C. L. Antisense transcription of a murine FGFR-3 pseudogene during fetal development. *Gene* **187**, 115–122 (1997).
24. Zhou, B. S., Beidler, D. R. & Cheng, Y. C. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. *Cancer Res.* **52**, 4280–4285 (1992).
25. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* doi:10.1038/nature06908 (this issue).
26. Stein, P., Zeng, F., Pan, H. & Schultz, R. M. Absence of non-specific effects of RNA interference triggered by long double-stranded RNA in mouse oocytes. *Dev. Biol.* **286**, 464–471 (2005).
27. Bettegowda, A. & Smith, G. W. Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. *Front. Biosci.* **12**, 3713–3726 (2007).
28. Stitzel, M. L. & Seydoux, G. Regulation of the oocyte-to-zygote transition. *Science* **316**, 407–408 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Hannon laboratory for discussions. O.H.T. is a Bristol-Meyers Squibb fellow and A.G. is a Florence Gould Foundation Scholar of the Watson School of Biological Sciences. E.P.M. is supported by a fellowship from the Australian-American Association. This work was supported in part by grants from the NIH (R.M.S. and G.J.H.) and gifts from Kathryn W. Davis and the Stanley family (G.J.H. and E.H.). G.J.H. is an Investigator of the Howard Hughes Medical Institute.

**Author Information** Small RNA data sets can be accessed in GEO with the following accession numbers GSM261957, GSM261958 and GSM261959. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.J.H. ([hannon@cshl.edu](mailto:hannon@cshl.edu)).



## METHODS

**Mouse strains.** Either CF-1 or CD-1 wild-type mice of 4–6 weeks were purchased from Harlan or Charles River Laboratories, respectively, and used to obtain oocytes for small RNA isolation. The *Dicer<sup>flox/flox</sup>* and *Dicer<sup>flox/flox</sup> Zp3-cre* mice, as previously reported<sup>14</sup>, were used to obtain Dicer-deficient oocytes.

**Generation of oocyte small RNA libraries.** Wild-type mice were primed with 5 IU (international unit) PMSG 48 h before being killed, and fully grown germinal vesicle oocytes were collected as previously described<sup>29</sup>. Total RNA was extracted using Trizol (Invitrogen) according to the manufacturer's protocol, and small RNA cloning was performed as described<sup>4</sup>.

**Quantitative real-time PCR.** Total RNA was extracted from fully grown oocytes from *Dicer<sup>flox/flox</sup>* and *Dicer<sup>flox/flox</sup> Zp3-cre* mice using the Absolutely RNA Microprep Kit (Stratagene). cDNA was prepared by reverse transcription of total RNA with Superscript II and random hexamer primers. One oocyte equivalent of the resulting cDNA was amplified using TaqMan probes and the ABI Prism Sequence Detection System 7000 (Applied Biosystems). Three replicates of 45 oocytes each were used for RNA isolation and two replicates were run for each real-time PCR reaction; a minus template served as control. Quantification was normalized to the endogenous upstream binding factor (*Ubf*) within the log-linear phase of the amplification curve obtained for each probe/primer using the comparative  $C_T$  method (ABI PRISM 7700 Sequence Detection System, User Bulletin 2, Applied Biosystems, 1997). The TaqMan gene expression assays used were: Mm00441071\_m1 (*Rangap1*), Mm00835842\_g1 (*Kifc1*), Mm00620601\_m1 (*Oog4*), Mm00786153\_s1 (*Lcp1*), Mm00728630\_s1 (*Kif2c*), Mm02391771\_g1 (*Hdac1*), Mm00487521\_m1 (*Mad1l1*), Mm00725286\_m1 (*Optn*), Mm00833431\_g1 (*Hsp90ab1*), Mm00511698\_m1 (*Ppp2r2b*), Mm00801709\_m1 (*Emp2*), Mm00486494\_m1 (*Surf6*), Mm00456972\_m1 (*Ubf*). For *Bcn1* and *Ubc9*, custom TaqMan Gene Expression Assays were used that had the following primers and probes: *Bcn1* forward primer 5'-ACTGGACGCTTCAGGATTACATC-3', *Bcn1* reverse primer 5'-GTCATGATGCTCCAGTGATCCA-3', *Bcn1* probe 5'-FAM-TTCCCAGAGGCATCCTG-3'; *Ubc9* forward primer 5'-CAGGTGAGAGCC-AAGGACAAA-3', *Ubc9* reverse primer 5'-GGCCCACTGTACAGCTAACA-3', *Ubc9* probe 5'-FAM-CTGGCCTGCATTGATC-3'.

**Bioinformatic analysis.** Small RNAs were sequenced using the Illumina 1G platform. Sequencing of the upper and lower fraction libraries produced 2,785,080 reads, of which 1,037,355 (37%) could be mapped to the mouse genome (release mm9, July 2007) with no mismatches. The small RNAs are matched to a suffix array generated from the mouse genome, keeping track of

exact matches to genome. Repeat masking was not used, but small RNA sequences with more than ten identical nucleotides in a row were removed from consideration. Annotation categories were assigned based on the annotation of corresponding genomic sequences extracted from the UCSC genome browser. The genome was annotated with mRNAs, non-coding RNAs and repeats. The annotations at the mapping positions (up to five mappings per small RNA) were used, along with a majority rule, to assign an annotation to each small RNA. The sequences were also re-analysed to allow 1–2 mismatches to the genome. Although the number of mapped sequences increased (from 37% to 54%), the genomic origin of repeat-associated small RNAs became ambiguous (Supplementary Fig. 1), and therefore non-informative.

To extract small RNA clusters (both piRNA and siRNA), the genome was scanned to look for regions that had more than ten uniquely mapping small RNAs, and the boundaries were defined as the location of the first/last small RNA in the cluster. To identify sequences that match the consensus for transposable elements, the small RNAs were aligned to consensus sequences from release 11.08 of Repbase (<http://www.girinst.org>). The following consensus sequences were used: L1\_MM for LINE L1 and IAPLTR1a\_I\_MM for the IAP retrotransposon. Matches to consensus sequences with up to three mismatches were recovered and included in the analysis. Nucleotide biases were calculated for small RNAs matching L1 and IAP consensus as described<sup>4</sup>. To identify gene–pseudogene pairs, the genomic sequences of the siRNA clusters were extracted from the UCSC genome browser, and re-matched to the genome using BLAT (<http://www.genome.ucsc.edu>). Genomic regions with greater than 95% identity were identified, and small RNAs (both sense and antisense) mapped to these locations were extracted.

Gene ontology analysis of endo-siRNA targets was carried out as previously described<sup>30</sup> using GOBAR, which uses a hypergeometric statistic to identify nodes that are significantly enriched. A bootstrapping technique, involving repeated sampling from the reference set, is used to assign significance values to the results.

29. Schultz, R. M., Montgomery, R. R. & Belanoff, J. R. Regulation of mouse oocyte meiotic maturation: implication of a decrease in oocyte cAMP and protein dephosphorylation in commitment to resume meiosis. *Dev. Biol.* **97**, 264–273 (1983).
30. Lee, J. S., Katari, G. & Sachidanandam, R. GObar: a gene ontology based analysis and visualization tool for gene sets. *BMC Bioinformatics* **6**, 189 (2005).

# Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes

Toshiaki Watanabe<sup>1,2</sup>, Yasushi Totoki<sup>3,†</sup>, Atsushi Toyoda<sup>4</sup>, Masahiro Kaneda<sup>5,6</sup>, Satomi Kuramochi-Miyagawa<sup>7</sup>, Yayoi Obata<sup>8</sup>, Hatsune Chiba<sup>1,2</sup>, Yuji Kohara<sup>2,9</sup>, Tomohiro Kono<sup>8</sup>, Toru Nakano<sup>7</sup>, M. Azim Surani<sup>8</sup>, Yoshiyuki Sakaki<sup>3,4</sup> & Hiroyuki Sasaki<sup>1,2</sup>

RNA interference (RNAi) is a mechanism by which double-stranded RNAs (dsRNAs) suppress specific transcripts in a sequence-dependent manner. dsRNAs are processed by Dicer to 21–24-nucleotide small interfering RNAs (siRNAs) and then incorporated into the argonaute (Ago) proteins<sup>1–4</sup>. Gene regulation by endogenous siRNAs has been observed only in organisms possessing RNA-dependent RNA polymerase (RdRP)<sup>5–10</sup>. In mammals, where no RdRP activity has been found, biogenesis and function of endogenous siRNAs remain largely unknown. Here we show, using mouse oocytes, that endogenous siRNAs are derived from naturally occurring dsRNAs and have roles in the regulation of gene expression. By means of deep sequencing, we identify a large number of both ~25–27-nucleotide Piwi-interacting RNAs (piRNAs) and ~21-nucleotide siRNAs corresponding to messenger RNAs or retrotransposons in growing oocytes. piRNAs are bound to Mili and have a role in the regulation of retrotransposons. siRNAs are exclusively mapped to retrotransposons or other genomic regions that produce transcripts capable of forming dsRNA structures. Inverted repeat structures, bidirectional transcription and antisense transcripts from various loci are sources of the dsRNAs. Some precursor transcripts of siRNAs are derived from expressed pseudogenes, indicating that one role of pseudogenes is to adjust the level of the founding source mRNA through RNAi. Loss of Dicer or Ago2 results in decreased levels of siRNAs and increased levels of retrotransposon and protein-coding transcripts complementary to the siRNAs. Thus, the RNAi pathway regulates both protein-coding transcripts and retrotransposons in mouse oocytes. Our results reveal a role for endogenous siRNAs in mammalian oocytes and show that organisms lacking RdRP activity can produce functional endogenous siRNAs from naturally occurring dsRNAs.

A large proportion of the mammalian genome produces both sense and antisense transcripts<sup>11–14</sup>. Although some models—such as transcriptional interference, RNA masking, RNA editing and RNAi—have been proposed<sup>15</sup>, the mechanisms underlying gene regulation by the diverse set of antisense RNAs remain largely unknown. In mammals, endogenous siRNAs derived from transposable elements have been identified in fully grown mouse oocytes and cultured human cells<sup>16,17</sup>; however, the number of siRNA molecules identified so far is small<sup>16</sup>, and their function and biogenesis remain unclear. To obtain a comprehensive picture of endogenous siRNAs,

we have sequenced more than 100,000 small RNAs from growing mouse oocytes.

The length distribution of the total small RNAs showed a bimodal pattern (Fig. 1a): one peak was observed at 21 nucleotides, corresponding to the length of microRNAs (miRNAs) and siRNAs, and the other at 25–26 nucleotides, corresponding to the length of piRNAs, which are a distinct class of small RNAs bound to Piwi family proteins<sup>16,18–23</sup>. Annotation of the small RNAs revealed that both the 21-nucleotide and 25–26-nucleotide small RNAs were mainly derived from repeat sequences, most of which were retrotransposons (Supplementary Fig. 2a and Supplementary Tables 1 and 2) and were quite diverse in sequence. We identified 21,969 clones of 21-nucleotide small RNAs comprising 10,194 different sequences and 18,572 clones of 25–26-nucleotide small RNAs comprising 12,006 different sequences.

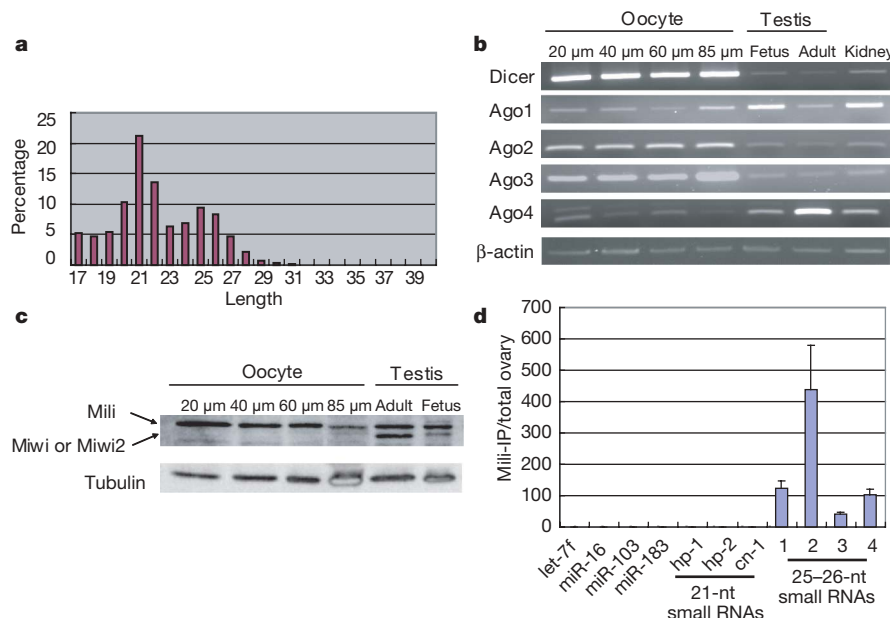
To examine expression of some components of the small RNA pathways in growing oocytes, we carried out polymerase chain reaction with reverse transcription (RT-PCR) and western blotting. Ago2, Ago3 and Dicer, which are the components of the siRNA and miRNA pathways, were expressed at high levels throughout oocyte growth (Fig. 1b). Of the three mouse Piwi family proteins, Mili was predominantly expressed at early stages of oocyte growth (Fig. 1c). However, we did not detect Miwi or Miwi2 in growing oocytes. To examine whether the 25–26-nucleotide small RNAs in growing oocytes are piRNAs bound to Mili, a total small RNA library and a Mili-immunoprecipitated (IP) small RNA library were constructed from ovaries, and the abundance of several 25–26-nucleotide small RNAs were examined in each library. All 25–26-nucleotide small RNAs were enriched in the Mili-IP library relative to the control miRNAs and 21-nucleotide small RNAs (Fig. 1d), indicating that the 25–26-nucleotide small RNAs in oocytes are mostly Mili-bound piRNAs.

Genomic mapping of the oocyte small RNAs (excluding those with more than ten hits to the genome, which are likely to be repeat sequences) revealed the presence of a total of 444 small RNA clusters (Supplementary Tables 3 and 4). piRNAs are known to be mapped to the genome in clusters<sup>16,18–23</sup>, and we identified 152 piRNA clusters by sequencing and mapping of Mili-IP small RNAs (Supplementary Fig. 2b and Supplementary Table 3). Notably, many of the largest clusters, which were determined by the number of small RNAs, were not included in these piRNA clusters (Supplementary Fig. 3a).

<sup>1</sup>Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan.

<sup>2</sup>Department of Genetics, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Mishima 411-8540, Japan. <sup>3</sup>Genome Annotation and Comparative Analysis Team, Computational and Experimental Systems Biology Group, and <sup>4</sup>Sequence Technology Team, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan.

<sup>5</sup>Wellcome Trust/Cancer Research UK Gurdon Institute of Cancer and Developmental Biology, University of Cambridge, Cambridge CB2 1QN, UK. <sup>6</sup>Reproductive Biology and Technology Research Team, National Institute of Livestock and Grassland Science, National Agriculture and Food Research Organization, Tsukuba 305-0901, Japan. <sup>7</sup>Department of Pathology, Graduate School of Medicine and Frontier Biosciences, Osaka University, Osaka 565-0871, Japan. <sup>8</sup>Department of BioScience, Tokyo University of Agriculture, Tokyo 156-8502, Japan. <sup>9</sup>Genome Biology Laboratory, Center for Genetic Resource Information, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan. <sup>†</sup>Present address: MetaSystems Research Team, RIKEN Advanced Science Institute, Yokohama 230-0045, Japan.

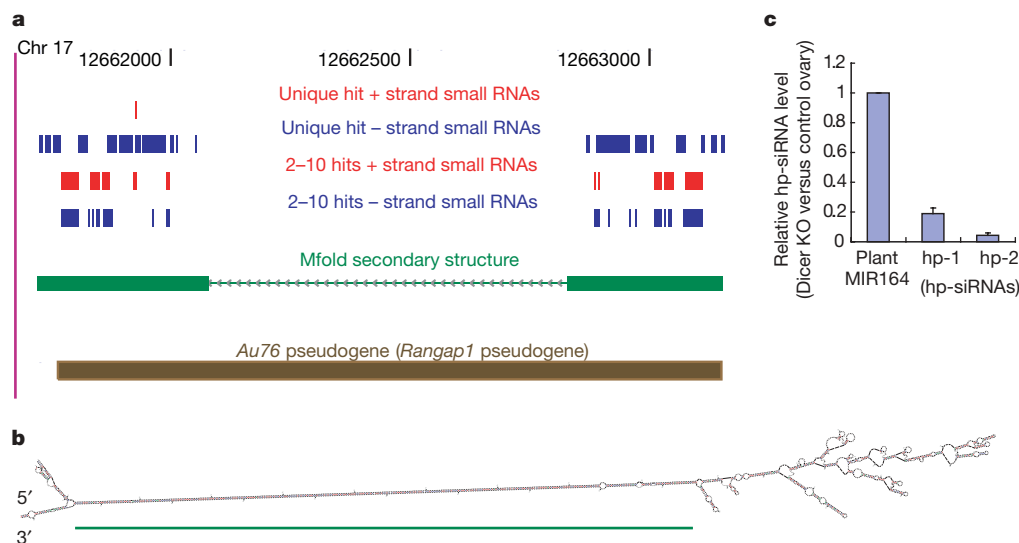


**Figure 1 | Small RNA profile and expression of small RNA pathway components in mouse oocytes.** **a**, Length distribution of total small RNAs (103,995 sequences) from oocytes. **b**, RT-PCR analysis of genes involved in the small RNA pathway. Oocytes with the diameter of ~20, ~40, ~60 and ~85  $\mu\text{m}$  (fully grown oocytes) were analysed. For comparison, mRNAs in testis and kidney were also analysed. **c**, Western blot analysis of Piwi family proteins in growing oocytes. The antibody used here detects all three Piwi family proteins. Positions of these proteins on the membrane are indicated on the left. The membrane was reprobed with the antibody against tubulin as

a loading control. **d**, Quantitative RT-PCR analysis of selected 21-nucleotide and 25–26-nucleotide small RNAs in total and Mili-IP small RNAs from P8 ovaries. The selected 21-nucleotide small RNAs are derived from hp-siRNA clusters (hp-1/2) and cis-nat-siRNA clusters (cn-1) (see below). For each RNA species tested, the amount in the Mili-IP library was divided by that in the total small RNA library. Error bars represent standard error ( $n = 3$ ). Some miRNAs were also tested. Amplified products were sequenced and confirmed to have the correct sequences.

Furthermore, the length distribution of the small RNAs constituting these non-piRNA clusters was centred at 21 nucleotides (Supplementary Fig. 3b), which is the length of miRNAs and siRNAs. However, the lack of a short stem-loop structure (Supplementary Fig. 4), which is a characteristic of miRNA precursors, in most of the genomic sequences encompassing the 21-nucleotide small RNAs indicated that they are not miRNAs.

The largest novel cluster was located at the *Au76* locus, a pseudogene of *Rangap1*. In this cluster, 979 clones of oocyte small RNAs comprising 485 different sequences were mapped in a 1,447-nucleotide region (Fig. 2a and Supplementary Table 3). Most (91%) of them were 19–22 nucleotides in length. Close inspection of this region revealed an inverted repeat structure in the *Au76* pseudogene (Fig. 2a, b). The small RNAs were exclusively mapped



**Figure 2 | Structure of the hp-siRNA cluster at the *Au76* locus.** **a**, An hp-siRNA cluster at the *Au76* locus on chromosome 17. The small RNAs mapped in this region are represented by red (plus strand) or blue (minus strand) bars. Small RNAs with unique hit and 2–10 time hits to the genome were indicated in different lines. The thick green bars in Mfold secondary structure represent the portions constituting the stem of the hairpin structure represented in **b** (green bar in **b**). *Au76*, a pseudogene of *Rangap1*, is indicated in brown. **b**, Secondary structure of the putative transcript derived from *Au76*. The

minus strand of the genomic sequence was folded using Mfold. The most stable structure predicted by Mfold is shown. **c**, Quantitative RT-PCR analysis of two hp-siRNAs derived from this locus in conditional Dicer knockout ovaries. The amounts of these RNAs in conditional Dicer knockout ovaries relative to those in ovaries that do not express Cre recombinase are shown. To both knockout and control samples, the same relative amount of plant MIR164 was added, which served as an external control. Error bars represent s.d. ( $n = 3$ ). Amplified products were sequenced and confirmed.



to this inverted repeat structure and were orientated in the same direction. Combined with the fact that production of these small RNAs requires Dicer and Ago2 (see below), these observations indicate that the small RNAs mapped to the *Au76* pseudogene locus are siRNAs that are produced from a precursor RNA with an intramolecular dsRNA structure (Fig. 2b). We designate this type of siRNA cluster as a 'hairpin siRNA (hp-siRNA) cluster'. Using our in-house program that detects inverted repeat structure in the small RNA clusters (see Methods), we identified three other hp-siRNA clusters in the mouse genome (Supplementary Table 3).

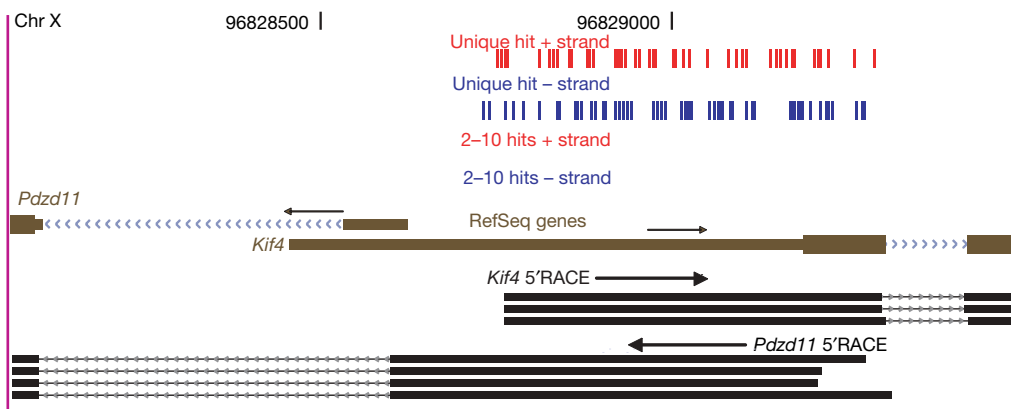
Only one mammalian Dicer protein has been identified and shown to be involved in the miRNA pathway<sup>24</sup>. The abundance of siRNAs derived from the *Au76* pseudogene was markedly decreased in conditional Dicer knockout ovaries (Fig. 2c), suggesting that Dicer is also involved in the production of siRNAs from intramolecular dsRNA precursors. Regulation of the founding source gene by a pseudogene has been reported in mammalian cells, but the existence of such regulation has been controversial<sup>25,26</sup>. The abundance of *Rangap1* mRNA, which is the founding source gene of *Au76* and shows ~90% nucleotide identity, was increased ~4-fold in Dicer knockout oocytes (Supplementary Fig. 5a), suggesting that *Au76* negatively regulates the founding source gene in *trans* through an RNAi mechanism.

Identification of hp-siRNA clusters led us to ask whether other types of siRNA clusters were present. Formation of dsRNAs can occur by transcription of natural antisense transcripts from the same loci or different loci. We designated such hypothetical siRNA clusters as 'cis-nat-siRNA clusters' and 'trans-nat-siRNA clusters', respectively. We searched for such clusters using our in-house program designed to detect the siRNA mapping patterns expected for these (Supplementary Methods). Seventeen loci met our criteria for cis-nat-siRNA clusters (Supplementary Table 3). An example of the predicted cis-nat-siRNA clusters was found at the *Pdzd11/Kif4* locus, where the two genes are orientated in a head-to-head manner (Fig. 3). At this locus, 135 clones of small RNAs comprising 117 different sequences were mapped to the first exon of the *Kif4* gene (Fig. 3). Of these small RNAs, 93% were 19–22 nucleotides in length. A 5' rapid amplification of cDNA ends (5'RACE) analysis of the transcripts from growing oocytes revealed that the first exons of *Pdzd11* and *Kif4* overlapped (Fig. 3). Notably, almost all of the small RNAs mapped to this locus were derived from this overlapping region, suggesting that these small RNAs were produced from an intermolecular dsRNA formed between the oppositely oriented transcripts. In Dicer mutants, levels of the siRNAs derived from this locus were decreased ~7-fold (Supplementary Fig. 5b), and both *Pdzd11* and *Kif4* mRNA levels were increased ~1.5-fold (Supplementary Fig. 5c), suggesting that the bidirectional overlapping transcription regulates *Pdzd11* and *Kif4* expression through RNAi.

A bioinformatics search predicted seven sets of trans-nat-siRNA clusters (Supplementary Table 3), all of which were pairs of an mRNA and its pseudogene. Of the seven pseudogenes, two were observed in the 3' untranslated region of unrelated mRNAs, and five were observed in intergenic regions. A representative trans-nat-siRNA cluster pair consisted of the *Ppp4r1* gene on chromosome 17 and its processed pseudogene on chromosome 8 (Fig. 4). *Ppp4r1* and its pseudogene showed ~90% nucleotide identity. At the *Ppp4r1* locus, the majority of 72 small RNAs (96% were 19–22 nucleotides in length) comprising 63 sequences mapped exclusively to the exons of *Ppp4r1*, and all unique small RNAs were orientated in the same direction as the gene (Fig. 4), suggesting that *Ppp4r1* mRNA was the source of the siRNAs. In the pseudogene locus, 77 small RNAs (88% were 19–22 nucleotides in length) comprising 69 sequences were mapped, and almost all unique small RNAs were orientated in the antisense direction of the *Ppp4r1* sequence (Fig. 4). Oocyte expressed-sequence-tags (ESTs) that mapped to this region were orientated in the same direction as the unique small RNAs, suggesting that the transcripts were the source of the siRNAs. We did not observe small RNAs in the 3' region of the last exon of *Ppp4r1* (right side in Fig. 4, top), even though it spanned approximately one-quarter of the total mRNA length. In the region of the pseudogene corresponding to this 3' region (left side in Fig. 4, bottom), no EST was observed. Thus, siRNAs were produced exclusively from the region where dsRNAs could be formed between the mRNA and its expressed pseudogene. These results are consistent with idea that dsRNAs were the only source of the siRNAs. In Dicer knockout oocytes, *Ppp4r1* mRNA level was increased ~1.5-fold (Supplementary Fig. 5d). Together, these results strongly suggest that the antisense transcripts from the *Ppp4r1* pseudogene suppress *Ppp4r1* expression through RNAi.

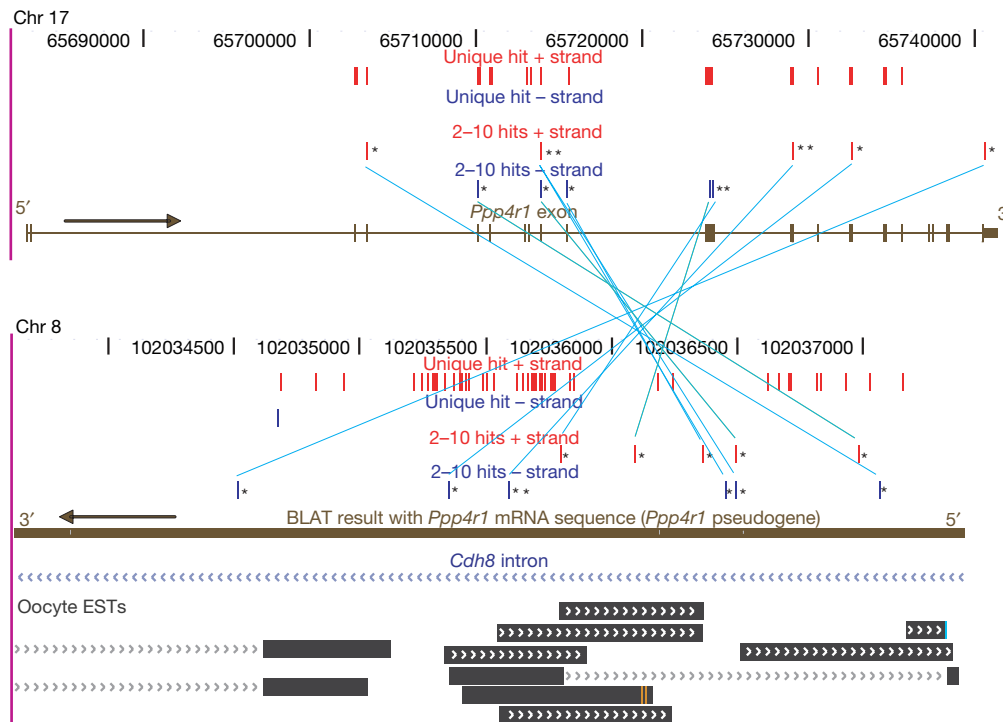
Most of the siRNAs and piRNAs in growing oocytes corresponded to retrotransposons. We therefore examined the possibility that the small RNA pathways suppress retrotransposons in oocytes. In Dicer knockout oocytes, we observed that the transcript level of retrotransposon LTR 10 (RLTR10) was elevated ~5-fold and that of mouse transcript A (MTA), which gives rise to more than 10% of the total polymerase II transcripts in mouse oocytes<sup>27</sup>, was elevated ~3-fold (Supplementary Fig. 6a). In Mili mutant oocytes, the transcript level of intracisternal-A particle (IAP) retrotransposons was elevated ~3.5-fold (Supplementary Fig. 6b). These data suggest that both piRNA and siRNA pathways suppress retrotransposons in mouse oocytes and that each pathway has preferred targets.

A peculiar siRNA cluster was observed in a retrotransposon-rich region. This ~50-kilobase (kb) locus contained 983 small RNAs comprising 637 different sequences (Supplementary Fig. 7). The locus produced both 25–26-nucleotide piRNAs and ~21-nucleotide small RNAs and was annotated as both a piRNA and an hp-siRNA



**Figure 3 | Structure of the cis-nat-siRNA cluster at the *Pdzd11/Kif4* locus.** A cis-nat-siRNA cluster at the *Pdzd11/Kif4* locus on chromosome X is shown. RefSeq gene predictions for *Pdzd11* and *Kif4* (rectangles, exons;

arrow head arrays, introns) are shown with the arrows indicating their transcriptional orientations. The transcribed regions determined by 5'RACE are indicated at the bottom (rectangles, exons; arrow head arrays, introns).



**Figure 4 | Structure of the trans-nat-siRNA cluster pair at the loci of *Ppp4r1* and its processed pseudogene.** A representative pair of trans-nat-siRNA clusters at the loci of *Ppp4r1* on chromosome 17 (top) and its processed pseudogene located in a *Cdh8* intron on chromosome 8 (bottom). The *Ppp4r1* exon/intron structure (top) and the processed pseudogene revealed by BLAT homology search (bottom) are indicated by brown bars.

Asterisks indicate the small RNAs mapped to both chromosome 8 and chromosome 17 clusters, and lines between the panels connect the locations of the same RNAs. Arrows indicate the 5' to 3' directions of *Ppp4r1* and its pseudogene. Black rectangles (bottom) represent oocyte ESTs, all of which are antisense to the *Ppp4r1* sequence.

cluster. A large fraction of the ~21-nucleotide small RNAs matched the RLTR10 sequence and were exclusively mapped to an ~2.5-kb inverted repeat structure located at the end of the piRNA cluster, suggesting that they were siRNAs produced from intramolecular dsRNAs. Overlaps between piRNA and siRNA clusters were also observed in other cases (7 out of 152 piRNA clusters and 36 siRNA clusters) (Supplementary Table 3).

To confirm that the siRNAs are assembled in the RNA-induced silencing complex (RISC), an Ago2-IP small RNA library and a total small RNA library were constructed from ovaries, and the abundance of some siRNAs were examined in each library. All siRNAs examined were enriched in the Ago2-IP library (Supplementary Fig. 8a), suggesting that these siRNAs are bound to Ago2. In Ago2 conditional knockout oocytes (M. Kaneda *et al.*, manuscript in preparation), the levels of mRNAs that are complementary to the siRNAs and those of retrotransposons that are elevated in Dicer knockout oocytes were increased (Supplementary Fig. 8b, c).

Our results, together with those by Tam *et al.*<sup>28</sup>, suggest that one role of pseudogenes may be to adjust the level of the founding source genes through RNAi. This kind of regulation may be conserved in other organisms but, as most of the pseudogenes that produce siRNAs were not found in the rat genome (data not shown), it may rather be used for species-specific fine tuning. In plants and *Caenorhabditis elegans*, accumulation of siRNAs requires RdRP activity<sup>5,7–10</sup>. By contrast, RdRP activity has not been found in mammalian cells. Our results provide evidence that regulation by endogenous siRNAs may be found in diverse organisms irrespective of the presence of RdRP activity.

## METHODS SUMMARY

**Small RNA library construction and sequencing.** To isolate small RNAs for library construction, 5 µg of total RNA from ~12,000 growing oocytes (35–60 µm) from B6D2F1 females and Mili-IP small RNAs from C57BL/6 ovaries at postnatal day 8 were used. Small RNA libraries were constructed using a Small

RNA Cloning Kit (Takara). For sequencing of the oocyte small RNA library, a 454 Life Sciences sequencer was used. Small RNAs were mapped to the genome using blastn. We used only perfect match sequences for further analysis. After annotation of the small RNAs, clusters were identified and classified using in-house programs.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 1 November 2007; accepted 10 March 2008.

Published online 10 April 2008.

- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Du, T. & Zamore, P. D. microPrimer: the biogenesis and function of microRNA. *Development* **132**, 4645–4652 (2005).
- Filipowicz, W., Jaskiewicz, L., Kolb, F. A. & Pillai, R. S. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* **15**, 331–341 (2005).
- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- Vaucheret, H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* **20**, 759–771 (2006).
- Ambros, V. & Chen, X. The regulation of genes and genomes by small RNAs. *Development* **134**, 1635–1641 (2007).
- Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241–244 (2007).
- Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
- Sijen, T., Steiner, F. A., Thijssen, K. L. & Plasterk, R. H. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**, 244–247 (2007).
- Yigit, E. *et al.* Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**, 747–757 (2006).
- Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
- Yelin, R. *et al.* Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnol.* **21**, 379–386 (2003).
- Chen, J. *et al.* Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res.* **32**, 4812–4820 (2004).
- Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).

15. Lavorgna, G. *et al.* In search of antisense. *Trends Biochem. Sci.* **29**, 88–94 (2004).
16. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20**, 1732–1743 (2006).
17. Yang, N. & Kazazian, H. H. Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13**, 763–771 (2006).
18. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).
19. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
20. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714 (2006).
21. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367 (2006).
22. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20**, 2214–2222 (2006).
23. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
24. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
25. Gray, T. A., Wilson, A., Fortin, P. J. & Nicholls, R. D. The putatively functional Mkrn1-p1 pseudogene is neither expressed nor imprinted, nor does it regulate its source gene in trans. *Proc. Natl Acad. Sci. USA* **103**, 12039–12044 (2006).
26. Hirotsune, S. *et al.* An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91–96 (2003).
27. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
28. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* doi:10.1038/nature06904 (this issue).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Tarakhovsky and D. O'Carroll for mice with the *Dicer* and *Ago2* conditional alleles; G. J. Hannon and A. Girard for the anti-AGO2 antibody; T. Sado, K. Hata and H. Furuumi for scientific and technical advice; Y. Kurihara, A. Takeda and K. Ichihara for comments on the manuscript; N. Minami and Y. Hoki for expertise in mouse oocytes; and K. Takada and M. Kiyooka for technical assistance. We thank RIKEN for the Super Combined Cluster (RSCC) computational resources. We also thank members of the Sasaki laboratory for discussion and encouragement. T.W. is a research fellow of the Japan Society for the Promotion of Science. This work was supported in part by Grants-in-Aid for Scientific Research on Priority Area from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to H.S.

**Author Contributions** T.W. performed experiments and interpreted results; Y.T. generated computational programs and analysed data with T.W.; A.T., Y.S. and Y.K. were involved in small RNA sequencing; Y.O., H.C. and T.K. were involved in oocyte collection; M.K. and M.A.S. prepared the samples from conditional *Dicer* and *Ago2* knockout mice; S.K.-M. and T.N. provided samples from Mili knockout mouse; H.S. and T.W. designed the study and wrote the manuscript.

**Author Information** The GenBank accession numbers of the small RNAs mapped to the genome in clusters and piRNAs co-immunoprecipitated with Mili are AB334800–AB349184 and AB349185–AB353040, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.W. ([toshwata@lab.nig.ac.jp](mailto:toshwata@lab.nig.ac.jp)) or H.S. ([hisasaki@lab.nig.ac.jp](mailto:hisasaki@lab.nig.ac.jp)).



## METHODS

**RT-PCR analyses.** For expression analysis of small RNA pathway components, 100–1,000 oocytes were collected from C57BL/6 female mice at various developmental stages and their total RNA (~100 ng) was reverse transcribed using random primers. For expression analysis of retrotransposons in Mili knockout oocytes, about 100 ng of total RNAs from 40–60- $\mu$ m oocytes (500 oocytes) was used. For expression analysis of retrotransposons and protein-coding transcripts in *Zp-3* conditional Dicer knockout oocytes<sup>29</sup> or *Zp-3* conditional Ago2 knockout oocytes, about 15 ng of total RNAs from 60–80  $\mu$ m oocytes (25 oocytes) was used. Before isolation of total RNA from Dicer knockout or Ago2 knockout oocytes, 10 pg of EGFP mRNA per oocyte was added. The primer sequences are listed in Supplementary Methods.

**Immunoprecipitation of Mili-piRNA complex and Ago2-siRNA complex.** A whole-cell extract was prepared from P8 (postnatal day 8) ovaries of C57BL/6 female mice in a lysis buffer (20 mM HEPES pH 7.3, 150 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.1% NP-40, 1 $\times$  Roche-Complete). Cleared extract was incubated with anti-Mili antibody<sup>18</sup> or anti-Ago2 antibody<sup>19</sup> for 12 h at 4 °C. Using protein G Sepharose, Mili-piRNA-antibody or Ago2-small RNA-antibody complexes were collected, and then washed four times in a wash buffer (20 mM HEPES pH 7.3, 320 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.1% NP-40, 1 $\times$  Roche-Complete) for 15 min.

**Small RNA library construction and sequencing.** We used 5  $\mu$ g of total RNA from ~12,000 growing oocytes (35–60- $\mu$ m) from B6D2F1 female mice, and Mili-IP small RNAs from C57BL/6 P8 ovaries (see above). Small RNAs ranging 15 to 40 nucleotides in length were cloned using a Small RNA Cloning Kit (TAKARA). For sequencing of the oocyte small RNA library, a 454 Life Sciences sequencer was used. Mili-IP small RNAs were sequenced using capillary sequencers. After trimming of the adaptor sequences, inserts were mapped to the mouse genome (mm8 assembly, February 2006) using blastn. Sequencing of the growing oocyte small RNA library yielded 176,267 reads of 17–40-nucleotide small RNAs, 103,995 (59%) of which comprised 63,244 non-redundant sequences that mapped to the genome with perfect match. Because our 454 sequencing of random 120-bp segments from the human genome resulted in a 2% error rate (A.T., unpublished data), we estimate only ~60% ( $\{100 - 2/100\}^{25} = 0.603$ ) would be correctly sequenced when sequencing 25-nucleotide small RNAs. Capillary sequencing of the Mili-IP small RNA library produced 4,937 reads of 17–40-nucleotide small RNAs, 4,129 of which were mapped to the genome with perfect match. We analysed only these perfect match sequences.

**Cluster identification.** After annotation of small RNAs (Supplementary Methods), small RNA clusters were identified. We used only small RNAs that

hit the genome 1–10 times and were annotated as repeat, piRNAs, mRNAs or unknown (the latter two can also include siRNAs and piRNAs). First, we scanned the genome using a 10-kb window and extracted the windows that had more than five small RNAs. Any overlapping windows that fulfilled these criteria were combined. If the combined region had more than three unique hit small RNAs, the region was considered to be a cluster. The positions of the most 5' and 3' small RNAs were considered to be the boundaries of the cluster. In cases where the boundary of the next cluster was located within 100 kb, the two clusters were considered to be one cluster (because both clusters may have been derived from a single precursor).

If one or more Mili-IP small RNAs were mapped within or around 1 kb of the cluster described above, we classified this cluster as a piRNA cluster. For this analysis we used Mili-IP small RNAs that uniquely hit the genome.

Identification and classification of the siRNA clusters were done using our in-house programs. The details of the selection procedures are described in Supplementary Methods.

**Small RNA library for PCR.** Mili-IP small RNAs from C57BL/6 P8 ovaries, Ago2-IP small RNAs from C57BL/6 P8 ovaries, total small RNAs from C57BL/6 P8 ovaries, total small RNAs from conditional Dicer knockout P15 ovaries<sup>29</sup>, total small RNAs from control P15 ovaries without Cre recombinase, *Mili*<sup>-/-</sup> P8–11 ovaries<sup>30</sup> and *Mili*<sup>+/-</sup> P8–11 ovaries were used for the construction of small RNA libraries for PCR. A synthetic RNA of plant MIR164 was added to the total RNAs from Dicer knockout and control ovaries before gel fractionation. Small RNA libraries were constructed as described above and amplified. In each quantitative PCR reaction, 0.5 ng of the amplified library was used. Individual small RNAs were amplified with specific primers complementary to the 3' part of the respective small RNAs and a universal primer corresponding to the 5' linker. Amplified products were sequenced and the sequences of the 5' part of the small RNA were confirmed. Only confirmed small RNAs were analysed. For quantification, each experiment was repeated three times. The primers are listed in Supplementary Methods.

**5'RACE analyses.** To determine the 5' end of cis-nat-siRNA precursors, 5'RACE was performed by using a GeneRacer kit (Invitrogen) according to the manufacturer's instructions. About 200 ng of total RNA from 40–60- $\mu$ m oocytes (~1,000 oocytes) from C57BL/6 female mice was used. The primer sequences are described in Supplementary Methods.

29. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21**, 644–648 (2007).

30. Kuramochi-Miyagawa, S. *et al.* Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).

## LETTERS

# Transcriptome-wide noise controls lineage choice in mammalian progenitor cells

Hannah H. Chang<sup>1,2,3</sup>, Martin Hemberg<sup>4†</sup>, Mauricio Barahona<sup>4</sup>, Donald E. Ingber<sup>1,5</sup> & Sui Huang<sup>1†</sup>

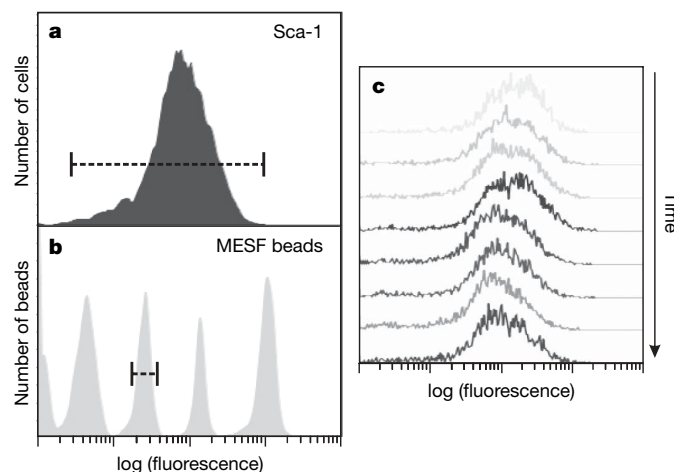
Phenotypic cell-to-cell variability within clonal populations may be a manifestation of ‘gene expression noise’<sup>1–6</sup>, or it may reflect stable phenotypic variants<sup>7</sup>. Such ‘non-genetic cell individuality’<sup>7</sup> can arise from the slow fluctuations of protein levels<sup>8</sup> in mammalian cells. These fluctuations produce persistent cell individuality, thereby rendering a clonal population heterogeneous. However, it remains unknown whether this heterogeneity may account for the stochasticity of cell fate decisions in stem cells. Here we show that in clonal populations of mouse haematopoietic progenitor cells, spontaneous ‘outlier’ cells with either extremely high or low expression levels of the stem cell marker Sca-1 (also known as Ly6a; ref. 9) reconstitute the parental distribution of Sca-1 but do so only after more than one week. This slow relaxation is described by a gaussian mixture model that incorporates noise-driven transitions between discrete subpopulations, suggesting hidden multi-stability within one cell type. Despite clonality, the Sca-1 outliers had distinct transcriptomes. Although their unique gene expression profiles eventually reverted to that of the median cells, revealing an attractor state, they lasted long enough to confer a greatly different proclivity for choosing either the erythroid or the myeloid lineage. Preference in lineage choice was associated with increased expression of lineage-specific transcription factors, such as a >200-fold increase in Gata1 (ref. 10) among the erythroid-prone cells, or a >15-fold increased PU.1 (Sfp1) (ref. 11) expression among myeloid-prone cells. Thus, clonal heterogeneity of gene expression level is not due to independent noise in the expression of individual genes, but reflects metastable states of a slowly fluctuating transcriptome that is distinct in individual cells and may govern the reversible, stochastic priming of multipotent progenitor cells in cell fate decision.

Cell-to-cell variability can be quantified by analysing the dispersion of expression levels of a phenotypic marker within a cell population. Flow cytometric analysis of EML cells, a multipotent mouse haematopoietic cell line<sup>12</sup>, revealed an approximately 1,000-fold range in the level of the constitutively expressed stem-cell-surface marker Sca-1 among individual cells within one newly derived clonal cell population (Fig. 1a). The heterogeneity of Sca-1 expression in this clonal population was highly consistent between measurements (Fig. 1c) and could not be attributed to measurement noise (Fig. 1b). Moreover, cell-cycle-dependent cell size variation contributed only 1% to the observed variability of Sca-1 levels per cell (Supplementary Discussion and Supplementary Fig. 1).

To characterize the dynamics by which population heterogeneity arises, cells with the highest, middle and lowest ~15% Sca-1 expression level (denoted henceforth as Sca-1<sup>low</sup>, Sca-1<sup>mid</sup> and Sca-1<sup>high</sup> fractions) were isolated from one clonal population using

fluorescence-activated cell sorting (FACS). Cells were stripped free of the staining antibody immediately after isolation and were cultured in standard growth medium. Within hours, all three fractions showed broadening of the narrow Sca-1 histograms obtained immediately after sorting (Fig. 2a), but more than 9 days elapsed before the three fractions regenerated Sca-1 histograms similar to that of the parental (unsorted) population (Fig. 2a). Therefore, the restoration of the wide range of Sca-1 surface-expression levels is a slow process (requiring more than 12 cell doublings) that is independent of initial Sca-1 expression levels. Clonal heterogeneity was also regenerated from subclones derived from randomly selected individual cells that had varying initial mean Sca-1 levels (Supplementary Fig. 2).

What drives the regeneration of the parental ‘bell-shaped’ histogram from the three sorted population fractions (Fig. 2a)? Although a variety of mechanisms may in principle underlie this behaviour (Supplementary Discussion and Supplementary Fig. 3 and 4), we consider here a general theoretical stochastic formulation. Because the genetic circuitry governing the expression of Sca-1 is poorly understood<sup>13</sup>, modelling the process explicitly with genetic circuits subjected to stochastic dynamics<sup>14</sup> is not feasible. Instead, we took a phenomenological approach to determine which general



**Figure 1 | Robust clonal heterogeneity.** **a, b**, Heterogeneity among clonal cells in Sca-1 protein expression, detected by immunofluorescence flow cytometry (**a**), was significantly larger than the resolution limit of flow cytometry approximated by measurement of reference fluorescent MESF<sup>24</sup> beads (**b**). The dashed lines show the difference in spread of the distributions as explained in the text. **c**, Stability of clonal heterogeneity in Sca-1 over three weeks.

<sup>1</sup>Vascular Biology Programme, Department of Pathology and Surgery, Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Programme in Biophysics, <sup>3</sup>MD-PhD Programme, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Bioengineering and Institute for Mathematical Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. <sup>5</sup>Harvard Institute for Biologically Inspired Engineering, Cambridge, Massachusetts 02139, USA. <sup>†</sup>Present addresses: Department of Ophthalmology, Children's Hospital Boston, Boston, Massachusetts 02215, USA (M.H.); Institute for Biocomplexity and Informatics, University of Calgary, Calgary, Alberta T2N 1N4, Canada (S.H.).

class of models of stochastic processes best describes the observed behaviour. The simplest model would be an elementary mean-reverting (Ornstein–Uhlenbeck) process<sup>15</sup> that includes both noise-driven diffusion (capturing the generation of cell–cell variability) and a drift towards the deterministic equilibrium (representing relaxation to the parental distribution mean; Supplementary Theoretical Methods). However, a simple Ornstein–Uhlenbeck process describes the data only poorly, because it fails to recapitulate accurately the growth of the long left tail (for example, 100-fold range for the Sca-1<sup>high</sup> fraction) in the histogram.

An alternative explanation is that the relaxation process is complicated by slow dynamics on a rugged potential landscape that consists of multiple quasi-discrete state transitions, the stochastic nature of which produces an additional source of variability<sup>16</sup>. Recent analysis of human myeloid progenitor cells has provided experimental evidence for the existence of multiple metastable states<sup>17</sup>, consistent with the dynamics of complex gene regulatory networks that control mammalian cell fates. We thus extended the simple Ornstein–Uhlenbeck model to include transitions between distinct states (virtual subpopulations) using a gaussian mixture model (GMM) as a first approximation to a multimodal system. As quantified by the Akaike information criterion (Supplementary Theoretical Methods), the data can be described by a minimal GMM model comprised of only two distinct states, each described as a gaussian, the parameters of which were obtained from the observed histograms in the stationary phase (time  $\geq 9$  days).

Our GMM model allowed us to partition cells in every measured histogram (time point) into two ‘virtual subpopulations’ (blue, subpopulation 1; red, subpopulation 2 in Fig. 2a) on the basis of the expression values of the individual cells, thus providing the time evolution of the mean  $\mu_i$  and the relative abundance (weight)  $w_i$  for each subpopulation  $i = 1, 2$  (Fig. 2b, c and Supplementary Theoretical Methods). This theoretical description suggests that the asymmetric broadening of the truncated histograms, as partially reflected in the changes in  $\mu$  for the two subpopulations (Fig. 2b), only accounts for a fraction of the restoration of the equilibrium heterogeneity. In contrast, stochastic transitions between the

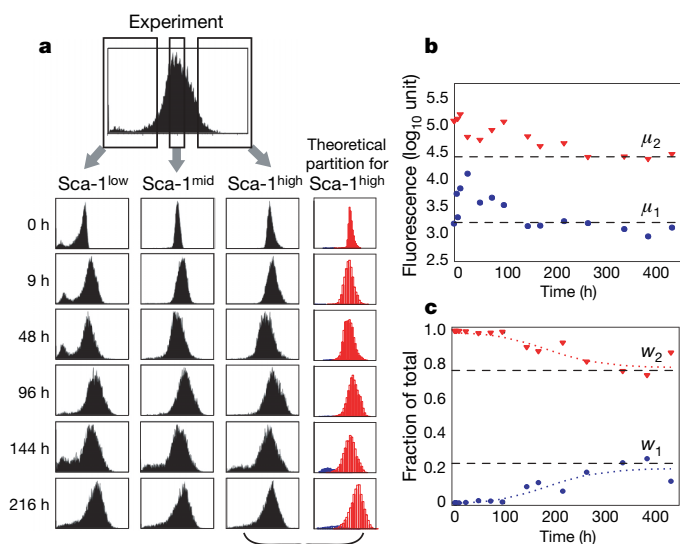
subpopulations, as reflected by the evolution of the weights  $w_i$ , had a dominant role in the later relaxation to equilibrium. Importantly, for the Sca-1<sup>mid</sup> and Sca-1<sup>high</sup> fractions, changes in  $w_i$  were initially negligible until 96 h, at which point the  $w_i$  exhibited a steep change before eventually reaching a plateau (Fig. 2c).

In summary, our results indicate that the observed clonal population heterogeneity of protein expression is not simply the manifestation of noise around a single, deterministic equilibrium (attractor) state described by an Ornstein–Uhlenbeck model. Instead, it is probably the result of processes involving stochastic state transitions in a system exhibiting multiple stable states<sup>17</sup>, which may explain the slow regeneration of the parental heterogeneity.

These results suggest that whole-population averaging of the level of Sca-1 may not appropriately characterize its biological function. Instead, owing to the slowness of relaxation to the mean values, momentary levels of Sca-1 within individual cells may reflect distinct, enduring functional states that have different biological consequences. Thus, we asked whether clonal heterogeneity in Sca-1 protein expression correlates with heterogeneity of the differentiation potential of these cells. Indeed, among the secondary clones generated from the parental population, the rate of commitment to pro-erythrocytes in response to erythropoietin (Methods and Supplementary Fig. 5) was inversely correlated to the baseline mean Sca-1 expression of each clone (Supplementary Fig. 6). Similarly, for the three sorted fractions (Fig. 3a), the relative erythroid differentiation rates were distinct, with Sca-1<sup>low</sup> cells differentiating the fastest, followed by Sca-1<sup>mid</sup> and Sca-1<sup>high</sup> (Fig. 3b). Importantly, although the Sca-1<sup>low</sup> fraction differentiated into the erythroid lineage at a rate sevenfold higher than the Sca-1<sup>high</sup> fraction (Fig. 3b), the Sca-1<sup>low</sup> fraction was not composed of spontaneously and irreversibly pre-committed pro-erythrocytes. Instead, these cells were still undifferentiated, as evidenced by expression of the stem cell marker c-kit (also known as Kit), their normal proliferation capacity (Supplementary Fig. 7) and their ability to reconstitute the parental histogram (Fig. 2a).

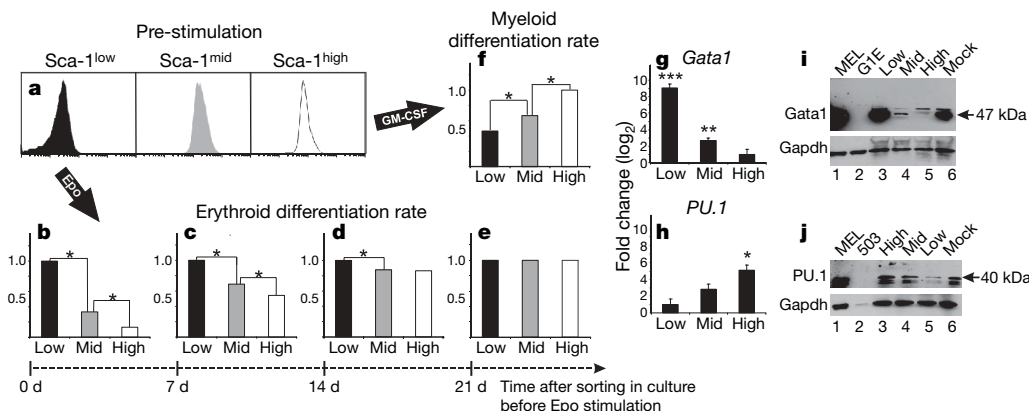
When we stimulated erythroid differentiation at various later time points after sorting, namely, on days 7, 14 and 21 of culture after sorting (as the Sca-1 histograms became more similar to each other while restoring the parental distribution), the difference in the erythroid differentiation rate between the Sca-1<sup>low</sup> and Sca-1<sup>high</sup> fractions was gradually lost (Fig. 3b–e). Surprisingly, despite the near complete convergence of the Sca-1 histograms at day 7, variability in differentiation kinetics was consistently detectable beyond 14 days after sorting (Fig. 3d). This suggests that clonal heterogeneity in Sca-1 expression controls differentiation potential but constitutes only a one-dimensional projection of separate states in the high-dimensional space of gene expression levels<sup>17</sup>. To reveal additional dimensions, we looked for correlated heterogeneity in other proteins and investigated whether expression of the erythroid-fate-determining transcription factor Gata1 (ref. 10) differed among the Sca-1 fractions. Real-time PCR revealed significantly higher *Gata1* messenger RNA levels in the erythroid differentiation-prone Sca-1<sup>low</sup> progenitor cells (260-fold increase over the Sca-1<sup>high</sup> fraction), followed by the Sca-1<sup>mid</sup> (2.7-fold increase over Sca-1<sup>high</sup> fraction) and Sca-1<sup>high</sup> fractions (Fig. 3g); these differences were paralleled by Gata1 protein levels (Fig. 3i). Importantly, *Gata1* mRNA expression among the three sorted fractions at 5 and 14 days after sorting (Supplementary Fig. 8) mirrored the gradual loss of variability observed in the differentiation kinetics for the erythroid lineage (Fig. 3b–e).

Gata1 has an antagonistic role to the myeloid-fate-determining transcription factor PU.1 in lineage determination; these two transcription factors mutually inhibit each other to regulate the erythroid versus myeloid fate decision<sup>18</sup>. Thus, we hypothesized that cells that are least prone to erythroid differentiation and exhibit low Gata1 expression may have high PU.1 levels, and thus be predisposed to the myeloid lineage. Indeed, real-time PCR revealed that Sca-1<sup>high</sup>



**Figure 2 | Restoration of heterogeneity from sorted cell fractions.** **a**, Clonal cells with the highest (Sca-1<sup>high</sup>), middle (Sca-1<sup>mid</sup>) and lowest (Sca-1<sup>low</sup>) 15% Sca-1 expression independently re-established the parental extent of clonal heterogeneity after 216 h in separate culture. As an example, each cell in the Sca-1<sup>high</sup> experiment was theoretically partitioned into one of two GMM subpopulations (blue and red, right). **b**, **c**, The temporal evolution of the means  $\mu_{1,2}$  (**b**) and weights  $w_{1,2}$  (**c**) for the Sca-1<sup>high</sup> GMM subpopulations 1 and 2. The evolution of the weights was fitted to a sigmoidal function (**c**, dotted curves). Black dashed lines, equilibrium values for  $\mu_i$  and  $w_i$ .

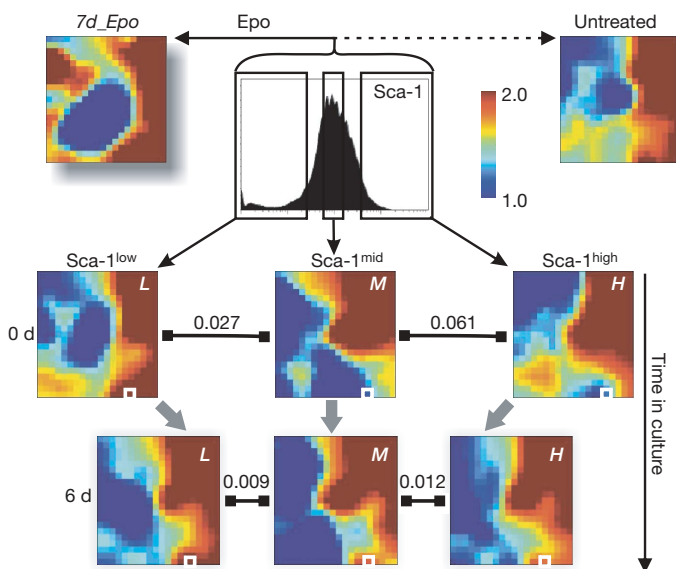




**Figure 3 | Clonal heterogeneity governs differentiation potential.** **a–f**, Sca-1<sup>low</sup> (Low, black), Sca-1<sup>mid</sup> (Mid, grey) and Sca-1<sup>high</sup> (High, white) fractions (**a**) stimulated by erythropoietin (Epo, **b**) and GM-CSF (**f**) immediately after isolation showed variable differentiation rates into the erythroid and myeloid lineages, respectively. After 7, 14 and 21 days (**d**) of post-sort culture, erythropoietin-treated cells showed convergence in both pre-stimulation, baseline Sca-1 expression (Fig. 2a) and relative differentiation rates (**b–e**). Asterisk,  $P < 0.001$  (two-tailed normal-theory test).

**g, h**, Quantitative real-time PCR with reverse transcription analysis of *Gata1* (**g**) and *PU.1* (**h**) mRNA levels in Sca-1-sorted fractions. Means  $\pm$  s.e.m. of triplicates shown. Triple asterisk,  $P < 10^{-5}$ ; double asterisk,  $P < 0.0002$ ; asterisk,  $P < 0.003$  (one-tailed Student's *t*-test). **i, j**, Western blot analysis of Gata1 (**i**) and PU.1 (**j**) protein levels in Sca-1 fractions (lanes 3–5) and mock-sorted cells (lane 6). The MEL cell line (lane 1) was used as a positive control. G1E and 503 (lane 2) cell lines were negative controls for Gata1 and PU.1, respectively. Gapdh was the loading control.

progenitor cells have the highest *PU.1* mRNA levels (17-fold increase over Sca-1<sup>low</sup> fraction), followed by the Sca-1<sup>mid</sup> (3.6-fold increase over Sca-1<sup>low</sup> fraction) and Sca-1<sup>high</sup> fractions (Fig. 3h). These differences were paralleled by PU.1 protein levels (Fig. 3j). Furthermore, myeloid differentiation rate was the highest among Sca-1<sup>high</sup> cells, followed by Sca-1<sup>mid</sup> and Sca-1<sup>low</sup> (Fig. 3f), in response to granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin 3 (IL-3; Methods and Supplementary Fig. 5). These results show that within a clonal population of multipotent progenitor cells, spontaneous non-genetic population heterogeneity primes the cells for different lineage choices.



**Figure 4 | Clonal heterogeneity of Sca-1 expression reflects transcriptome-wide noise.** Self-organizing maps of global gene expression for a subset of 2,997 genes visualized with the GEDI<sup>23</sup> program for Sca-1<sup>low</sup> (L), Sca-1<sup>mid</sup> (M), Sca-1<sup>high</sup> (H) fractions at 0 and 6 d after FACS isolation and for a differentiated erythroid culture (7 d erythropoietin, Epo) and an untreated control sample. Pixels in the same location within each GEDI map contain the same minicluster of genes. The colour of pixels indicates the centroid value of gene expression level for each minicluster in log<sub>10</sub> units of signal. Dissimilarity between transcriptomes is indicated above the horizontal distance symbols. The Gata1-containing pixel is boxed in white.

Because both Gata1 and PU.1 are pivotal lineage-specific transcription factors, we asked whether the marked upregulation of Gata1 and associated downregulation of PU.1 in the most erythroid-prone Sca-1<sup>low</sup> cells reflect a particular cellular state in terms of genome-wide gene expression. Microarray-based mRNA expression profiling on Sca-1<sup>low</sup> (L), Sca-1<sup>mid</sup> (M) and Sca-1<sup>high</sup> (H) fractions immediately after sorting revealed that these three fractions differed considerably in their transcriptomes (Fig. 4). Replicate microarray measurements showed that the observed transcriptome differences could not be attributed solely to experimental error (Supplementary Fig. 9). Significance analysis of microarrays (SAM)<sup>19</sup> revealed >3,900 genes that were differentially expressed between the Sca-1<sup>low</sup> and Sca-1<sup>high</sup> fractions at a stringent false detection rate of 1.5%. The distinct global gene expression profiles of the three fractions converged to a common pattern within 6 days after sorting, a progression that can be quantified by the inter-sample distance metric  $D = 1 - R$ , where  $R$  is the Pearson correlation coefficient. The distances between the three profiles decreased from  $D(L - M)_{0 \text{ days}} = 0.027$  to  $D(L - M)_{6 \text{ days}} = 0.009$  and from  $D(M - H)_{0 \text{ days}} = 0.061$  to  $D(M - H)_{6 \text{ days}} = 0.012$  (Fig. 4 and Supplementary Table 1). Thus, the outlier populations reconstituted the traits of the parental population not only with respect to their distribution of Sca-1 expression (Fig. 2a) and differentiation rates (Fig. 3b–e) but also with respect to their gene expression profiles across thousands of genes. This global relaxation from both ends of the parental spectrum towards the centre is predicted by the model in which a stable cell phenotype, such as the progenitor state here, is a high-dimensional attractor state<sup>20</sup>. It also confirms that the Sca-1 outlier cells were not already irreversibly committed. Nevertheless, Sca-1<sup>low</sup> cells exhibited a transcriptome that was clearly more similar than the Sca-1<sup>high</sup> cells to the unsorted but maximally differentiated cells, achieved by culture in erythropoietin for 7 days (7d\_Epo) (Fig. 4):  $D(L - 7d\_Epo) = 0.079$  versus  $D(H - 7d\_Epo) = 0.158$ ; Supplementary Table 1. This is a remarkable feat given the spontaneity and stochasticity of the process that generated these differentiation-prone outlier cells. In fact, with respect to 200 'differentiation marker genes' (Methods), only the Sca-1<sup>low</sup> cells were statistically similar to the erythropoietin-treated cells ( $P < 3 \times 10^{-14}$ , pairwise *t*-test), whereas the Sca-1<sup>mid</sup> ( $P > 0.8$ ) and Sca-1<sup>high</sup> ( $P > 0.6$ ) cells were not, further confirming the transcriptome similarity between the Sca-1<sup>low</sup> and erythropoietin-treated cells, which may be related to their increased Gata1 levels.

Our results demonstrate the robust nature of cell-to-cell variability that underlies the heterogeneity of gene expression in a clonal population of mammalian progenitor cells. Although the source of the heterogeneity and the molecular mechanisms responsible for its slow restoration remain to be elucidated, our experiments and general theoretical considerations point to discrete transitions in a dynamical system exhibiting multistability as one source of this behaviour. Independent of the specific mechanism, we show that biological function in metazoan cells is not necessarily determined by the ensemble average of a nominally homogenous cell population, and that outliers in a heterogeneous cell population do not simply represent irrelevant, short-lived phenotypic states caused by random fluctuations in the expression of a single gene. Instead, the departure from the average state is characterized by slowly fluctuating transcriptome-wide noise that has significant biological functionality in the priming of cell fate commitment. This finding helps unite two old dualisms: between plasticity and heterogeneity in explaining multipotency<sup>21,22</sup>, and between instructive and selective regulation in explaining cell fate decisions<sup>18</sup>. Exploiting the spontaneous and transient yet enduring cell individuality in differentiation potential resulting from clonal heterogeneity also could be of practical value in attempts to steer lineage choice in stem cells for therapeutic applications.

## METHODS SUMMARY

**Creation of single-cell-derived subclones.** Single-cell-derived subclones of EML cells were generated in three weeks by methylcellulose-plating at low cell densities, isolation of resulting colonies by hand with microscopic guidance, and expansion in liquid culture.

**Flow cytometry and bead calibration.** Cell surface protein immunostaining and flow cytometry measurements were performed using standard methods. For cells that were recultured after FACS, the staining antibody was removed as previously reported<sup>17</sup>. Quantum PE molecules of equivalent soluble fluorochrome (MESF) beads (Bangs Laboratories) were used to correct for daily fluctuations in flow cytometer sensitivity.

**Gene expression profiling with microarrays.** The MouseWG-6v1.1 Illumina microbead chips were used to perform gene expression profiling on total RNA extracted from FACS-sorted, or unsorted, cell populations.

**Data analysis.** Flow cytometry data were analysed using the software package FlowJo 2.2.2. Theoretical modelling and filtering of microarray data were performed with custom software written in Matlab 7.2. Statistical significance analysis of the microarray data was performed with the SAM<sup>19</sup> algorithm and self-organizing maps generated with the gene expression dynamics inspector (GEDI) software<sup>23</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 January; accepted 31 March 2008.

1. Blake, W. J. *et al.* Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
2. Elowitz, M. B. *et al.* Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
3. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
4. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
5. Rosenfeld, N. *et al.* Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).

6. Kaern, M. *et al.* Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
7. Spudich, J. L. & Koshland, D. E. Jr. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
8. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
9. van de Rijn, M. *et al.* Mouse hematopoietic stem-cell antigen Sca-1 is a member of the Ly-6 antigen family. *Proc. Natl Acad. Sci. USA* **86**, 4634–4638 (1989).
10. Cantor, A. B., Katz, S. G. & Orkin, S. H. Distinct domains of the GATA-1 cofactor FOG-1 differentially influence erythroid versus megakaryocytic maturation. *Mol. Cell. Biol.* **22**, 4268–4279 (2002).
11. Koschmieder, S. *et al.* Role of transcription factors C/EBP $\alpha$  and PU.1 in normal hematopoiesis and leukemia. *Int. J. Hematol.* **81**, 368–377 (2005).
12. Tsai, S. *et al.* Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. *Genes Dev.* **8**, 2831–2841 (1994).
13. Holmes, C. & Stanford, W. L. Concise review: stem cell antigen-1: expression, function, and enigma. *Stem Cells* **25**, 1339–1347 (2007).
14. Guido, N. J. *et al.* A bottom-up approach to gene regulation. *Nature* **439**, 856–860 (2006).
15. Uhlenbeck, G. E. & Ornstein, L. S. On the theory of Brownian Motion. *Phys. Rev.* **36**, 823–841 (1930).
16. Kurchan, J. & Laloux, L. Phase space geometry and slow dynamics. *J. Phys. Math. Gen.* **29**, 1929–1948 (1996).
17. Chang, H. H. *et al.* Multistable and multistep dynamics in neutrophil differentiation. *BMC Cell Biol.* **7**, 11 (2006).
18. Huang, S. *et al.* Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).
19. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
20. Huang, S. *et al.* Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701 (2005).
21. Enver, T., Heyworth, C. M. & Dexter, T. M. Do stem cells play dice? *Blood* **92**, 348–351; discussion 352 (1998).
22. Orkin, S. H. & Zon, L. I. Hematopoiesis and stem cells: plasticity versus developmental heterogeneity. *Nature Immunol.* **3**, 323–328 (2002).
23. Eichler, G. S., Huang, S. & Ingber, D. E. Gene expression dynamics inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* **19**, 2321–2322 (2003).
24. Zenger, V. E. *et al.* Quantitative flow cytometry: inter-laboratory variation. *Cytometry* **33**, 138–145 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by grants to S.H. from the Air Force Office of Scientific Research and, in part, from the National Institutes of Health. H.H.C. is partially supported by the Presidential Scholarship and the Ashford Fellowship of Harvard University. M.H. and M.B. are supported by the Life Sciences Interface and Mathematics panels of the Engineering and Physical Sciences Research Council of the UK. D.E.I. is supported by the National Health Institutes and the Army Research Office. We thank K. Orford, P. Zhang, A. Mammoto, J. Daley, J. Pendse and M. Shakya for experimental assistance, and W. Press and K. Farh for discussions.

**Author Contributions** H.H.C. designed the study, performed the experiments, analysed the data, participated in the theoretical analysis and drafted the manuscript. M.H. constructed the theoretical model and performed the theoretical analysis. M.B. constructed the model, supervised the work and revised the manuscript. D.E.I. supervised the work and revised the manuscript. S.H. conceived of the study, designed experiments, supervised the work, participated in the experimental and theoretical analysis and drafted the manuscript. All authors read and approved the final manuscript.

**Author Information** The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) under the GEO Series accession number GSE10772. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.H. ([sui.huang@ucalgary.ca](mailto:sui.huang@ucalgary.ca)).

## METHODS

**Culture of EML cells, derivation of subclones, and differentiation.** EML cells<sup>26</sup> (a gift from K. Orford and D. Scadden) were maintained in growth medium containing Iscove's modified Dulbecco's medium (IMDM), 20% horse serum, 12–15% (v/v) medium conditioned (CM) by baby hamster kidney (BHK) cells producing murine kit-ligand (MKL), and 1% glutamine/penicillin/streptomycin. To obtain single-cell-derived subclones, cells were plated into 60-mm plates at 500–2,000 cells ml<sup>-1</sup> density in 1% methylcellulose (Methocult M3134) containing growth medium and incubated without disturbance for 10 days. Individual well-demarcated colonies were hand-picked with Pasteur pipettes under microscopic guidance and were transferred to liquid cultures in microwell plates. Typical subclones required ~18 days in culture to expand to a sufficiently large population for the experiment. To differentiate EML cells into the erythroid lineage, a previously reported differentiation protocol<sup>12</sup> was adapted. In brief, on day 1, cells were cultured in growth medium plus 10 ng ml<sup>-1</sup> mouse recombinant erythropoietin (Sigma-Aldrich) at 250,000 cells ml<sup>-1</sup> density. On day 3, cells were spun down and re-suspended into IMDM plus 20% horse serum, 2% BHK/MKL-CM and 10 ng ml<sup>-1</sup> mouse recombinant erythropoietin at 125,000 cells ml<sup>-1</sup> density to give resulting erythroid cells a growth advantage. One day 6, an additional 10 ng ml<sup>-1</sup> of erythropoietin was added. Typically, 7 days of erythropoietin treatment generated ~40–60% (of total) pro-erythrocytes that were benzidine-stain-positive and Sca-1/c-kit double-negative (Supplementary Fig. 5). Benzidine staining was performed following a reported protocol<sup>25</sup> and examined by microscopy after cytospin. To differentiate EML cells into myeloid cells, a previously reported differentiation protocol<sup>12</sup> was adapted. In brief, on day 1, cells were cultured in growth medium plus 10 ng ml<sup>-1</sup> mouse recombinant IL-3 (Peprotech) and 10<sup>-5</sup> M retinoic acid (Sigma-Aldrich) at 300,000 cells ml<sup>-1</sup> density. On day 4, cells were washed thoroughly with PBS to remove remaining SCF from the growth medium and were cultured in IMDM plus 20% horse serum, 2% BHK/MKL-CM, 10 ng ml<sup>-1</sup> mouse recombinant GM-CSF (R&D Systems) and 10<sup>-5</sup> M retinoic acid (Sigma-Aldrich) at 200,000 cells ml<sup>-1</sup> density. On day 6, an additional 10 ng ml<sup>-1</sup> GM-CSF was added. After 7–9 days, differentiated myeloid cells dominate the culture and show Mac-1 (Itgam, integrin  $\alpha$  M) and Gr-1 (Ly6G) expression by flow cytometry.

**Flow cytometry, FACS and bead calibration.** For direct cell-surface-protein immunostaining, the antibodies Sca-1-PE (Caltag) and c-kit-FITC (BD Pharmingen) were used at 1:1,000 dilutions in ice-cold PBS plus 1% fetal calf serum with (flow cytometry) or without (FACS) 0.01% NaN<sub>3</sub>. Appropriate isotype control antibodies (BD Pharmingen) were used to establish background signal caused by non-specific antibody binding. Propidium iodide staining was correlated with lower forward scatter among EML cells (Supplementary Fig. 10). Thus, dead cells with positive propidium iodide staining were easily removed from all analysis by gating out the low forward scatter population. Flow cytometry was performed on a Becton Dickinson FACSCaliber analyser and FACS with either a Becton Dickinson FACS Aria or an AriaSpecial Sorter ultraviolet laser system at the Dana Farber Cancer Institute Flow Cytometry Core.

Computational data analysis was done with FlowJo 2.2.2. For cell sorting, input cell number ranged from 60 × 10<sup>6</sup> cells to 100 × 10<sup>6</sup> cells. Cells were sorted into ice-cold medium for a maximal duration of 3 h. Gates for the lowest, middle and highest Sca-1 expressors were set based on the proportion of total population. For cells that were re-cultured after FACS, the staining antibody was removed following a previously reported protocol<sup>17</sup>. Quantum PE MESF beads (Bangs Laboratories) were used to correct for the effect of day-to-day fluctuations in the flow cytometer following the manufacturer's instructions. Calibration curves were constructed using Matlab 7.2 (MathWorks) and were used to convert obtained fluorescence data into absolute MESF units for the purpose of quantitative theoretical modelling.

**Gene expression profiling with microarrays and data analysis.** Gene expression profiling was performed at the Molecular Genetics Core facility at the Children's Hospital Boston using MouseWG-6 v1.1 microbead chips (Illumina). Raw gene expression data were first subjected to rank-invariant normalization using BeadStudio 3.0. Matlab 7.2 was used to filter the list of 46,628 genes on the basis of two sets of criteria. First, detection *P*-value based on Illumina replicate gene probes: genes with detection *P*-values > 0.01 in all samples were called 'absent' in all samples and thus removed (giving rise to set 1, consisting of 14,038 genes). Genes with differing 'detection call' ('absent' versus 'present') between the duplicate samples were also removed. Second, fold-change: genes that did not show at least a twofold change compared to the Sca-1<sup>mid</sup> fraction in 4 out of the 12 total samples were also removed (resulting in set 2: 2,997 genes). Alternatively, the SAM<sup>19</sup> algorithm was used to filter by fold change at a stringent false detection rate of 1.5% (resulting in set 3: 3,973 genes). Qualitative conclusions did not depend on the exact stringency of the filtering. After filtering, gene expression levels were transformed by log<sub>10</sub> and subjected to clustering analyses. GEDI maps for visual representation of global gene expression based on self-organizing maps were generated using the program GEDI<sup>23</sup> (<http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm>). In GEDI, each 'tile' within a 'mosaic' represents a miniclust of genes that have highly similar expression pattern across all the analysed samples. The same genes are forced to the same mosaic position for all GEDI maps, hence allowing direct comparison of transcriptomes based on the overall mosaic pattern. The colour of tiles indicates the centroid value of gene expression level for each miniclust. Dissimilarity between samples was quantified by 1 - *R*, where *R* is the Pearson's correlation coefficient calculated for all genes in a pair of samples. For statistical analysis of the similarity between the sorted fractions and the erythropoietin-treated sample, a subset of ~200 'differentiation marker genes' were obtained from stringent SAM analysis of the unsorted, untreated control and the unsorted, erythropoietin-treated sample.

25. Wang, R., Clark, R. & Bautch, V. L. Embryonic stem cell-derived cystic embryoid bodies form vascular channels: an *in vitro* model of blood vessel development. *Development* **114**, 303–316 (1992).
26. Tsai, S. *et al.* Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. *Genes Dev.* **8**, 2831–2841 (1994).



## LETTERS

# Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction

Patrick Schreiner<sup>1\*</sup>, Xiang Chen<sup>2\*</sup>, Koraljka Husnjak<sup>3,4\*</sup>, Leah Randles<sup>2</sup>, Naixia Zhang<sup>2</sup>, Suzanne Elsasser<sup>5</sup>, Daniel Finley<sup>5</sup>, Ivan Dikic<sup>3,4,6</sup>, Kylie J. Walters<sup>2</sup> & Michael Groll<sup>1,7</sup>

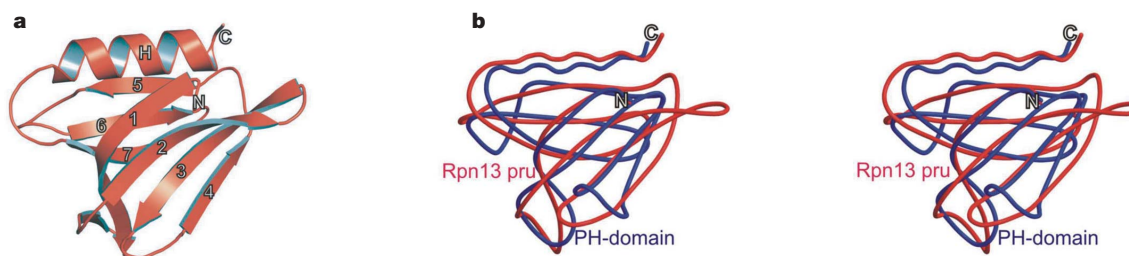
Targeted protein degradation is largely performed by the ubiquitin–proteasome pathway, in which substrate proteins are marked by covalently attached ubiquitin chains that mediate recognition by the proteasome. It is currently unclear how the proteasome recognizes its substrates, as the only established ubiquitin receptor intrinsic to the proteasome is Rpn10/S5a (ref. 1), which is not essential for ubiquitin-mediated protein degradation in budding yeast<sup>2</sup>. In the accompanying manuscript we report that Rpn13 (refs 3–7), a component of the nine-subunit proteasome base, functions as a ubiquitin receptor<sup>8</sup>, complementing its known role in docking de-ubiquitinating enzyme Uch37/UCHL5 (refs 4–6) to the proteasome. Here we merge crystallography and NMR data to describe the ubiquitin-binding mechanism of Rpn13. We determine the structure of Rpn13 alone and complexed with ubiquitin. The co-complex reveals a novel ubiquitin-binding mode in which loops rather than secondary structural elements are used to capture ubiquitin. Further support for the role of Rpn13 as a proteasomal ubiquitin receptor is demonstrated by its ability to bind ubiquitin and proteasome subunit Rpn2/S1 simultaneously. Finally, we provide a model structure of Rpn13 complexed to diubiquitin, which provides insights into how Rpn13 as a ubiquitin receptor is coupled to substrate deubiquitination by Uch37.

The structure of murine Rpn13 (mRpn13) (amino acids 1–150) was determined at 1.7-Å resolution by X-ray crystallography, and found to contain a pleckstrin-homology domain (PHD) fold (Fig. 1a, b) (structure determination and refinement statistics are provided in the Supplementary Information). In particular, whereas

the first 21 amino- and last 20 carboxy-terminal amino acids are unstructured, residues 22–130 form a PHD fold. This result was surprising, as primary sequence alignment did not identify Rpn13 as being homologous to previously characterized proteins. This finding, coupled with its ubiquitin receptor properties<sup>8</sup>, prompted us to name the N-terminal domain of Rpn13 pleckstrin-like receptor for ubiquitin (Pru).

Though very divergent at their sequence level, all PHDs have a common  $\beta$ -sandwich fold. The PHD of Rpn13 is composed of a four-stranded twisted antiparallel  $\beta$ -sheet ( $\beta_{1-4}$ : residues 22–34, 45–52, 56–62, 71–74) that packs almost orthogonally against a second triple-stranded  $\beta$ -sheet ( $\beta_{5-7}$ : residues 80–85, 92–98, 103–110) (Supplementary Fig. 1). Like other PHDs, Rpn13 Pru forms a hydrophobic core containing conserved hydrophobic residues (F26, V47, I49, F59, F82, Y94, L96, F107 and M109), which are located within  $\beta$ -sheets. One end of the  $\beta$ -sandwich is capped by a long C-terminal amphipathic  $\alpha$ -helix (residues 117–128), which is stabilized by interactions between V124 and L128, whereas the other corner of the hydrophobic core is closed by three loops formed by residues located between strands S1/S2, S3/S4 and S6/S7 (Fig. 1a and Supplementary Fig. 1).

Despite much effort, we were unable to crystallize the Rpn13 Pru–ubiquitin complex; however, we determined the structure of this complex by molecular docking, based on the crystal structure of mRpn13 Pru and intermolecular nuclear Overhauser enhancements (NOEs) as well as data from chemical shift perturbation derived by NMR titration experiments. The topology of the complexed structure was readily defined by 12 unambiguous intermolecular NOE



**Figure 1 | Crystal structure of mRpn13 Pru reveals typical pleckstrin-homology fold. a**, Ribbon representation of Rpn13 pleckstrin-like receptor for ubiquitin (Rpn13 Pru). The pleckstrin-homology fold consisting of a

seven-stranded  $\beta$ -sandwich structure (1–7) capped by the C-terminal  $\alpha$ -helix. **b**, Stereo representation of the structural alignment of Rpn13 Pru (red) and the PHD (blue) from Pleckstrin (PDB accession code 1PLS).

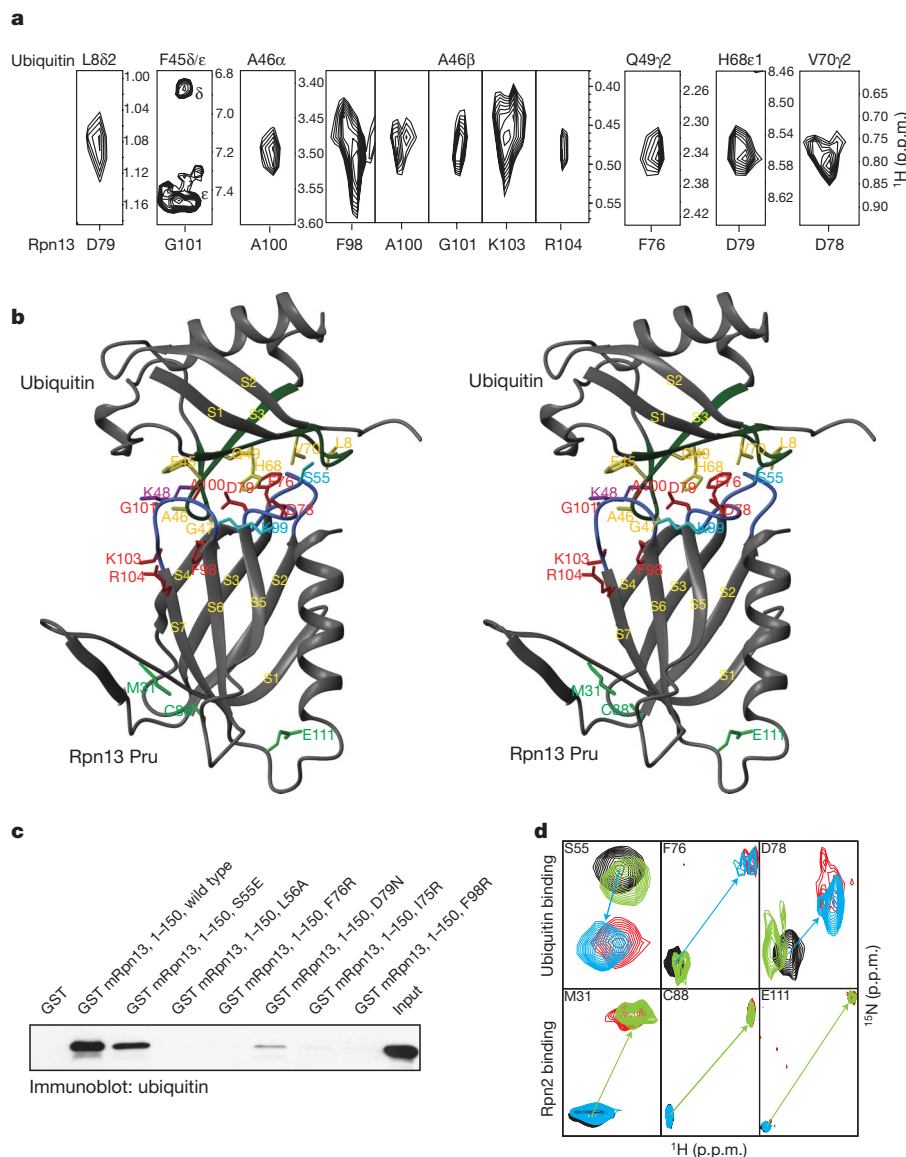
<sup>1</sup>Center for Integrated Protein Science at the Department Chemie, Lehrstuhl für Biochemie, Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany. <sup>2</sup>Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>3</sup>Institute for Biochemistry II and Cluster of Excellence Macromolecular Complexes, Goethe University, Theodor-Stern-Kai 7, D-60590 Frankfurt (Main), Germany. <sup>4</sup>Tumor Biology Program, Mediterranean Institute for Life Sciences, Mestroviccevo setaliste, 21000 Split, Croatia. <sup>5</sup>Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>6</sup>Department of Immunology, Medical School University of Split, Soltanska 2, 21000 Split, Croatia. <sup>7</sup>Institute of Biochemistry, Charité-Universitätsmedizin Berlin CCM, Monbijoustraße 2, D-10117 Berlin, Germany.

\*These authors contributed equally to this work.

interactions between human Rpn13 (hRpn13) Pru and ubiquitin (Fig. 2a). We were able to use the hRpn13:ubiquitin NOEs with the mRpn13 crystal structure because all of the amino acids exhibiting intermolecular NOEs are strictly conserved between murine and human Rpn13. Importantly, the mRpn13 Pru–ubiquitin structure reveals a novel ubiquitin-binding mode in which residues of the S2–S3, S4–S5 and S6–S7 loops capture ubiquitin (Fig. 2b). At the core of the contact surface, hydrogen bonds are formed between side-chain oxygens of D78 and D79 in hRpn13, and Nε2 and Nδ1 of H68 in ubiquitin, respectively. Moreover, F76 engages in hydrophobic interactions with I44, Q49 and V70 of ubiquitin (Supplementary

Fig. 2). These contacts are enabled by the strictly conserved P77, which causes the S4–S5 loop to turn. The intermolecular NOE data for this complex were fully satisfied without Rpn13 Pru or ubiquitin structural rearrangements, and the root mean squared deviations between the free and complexed state of Rpn13 Pru and ubiquitin were 0.91 and 0.75 Å for backbone atoms, and 1.20 and 1.15 Å for all non-hydrogen atoms, respectively.

Additional hydrophobic contacts exist, as Rpn13's side-chain methyl group of A100 and the Cα group of G101 partly bury ubiquitin's F45, which is solvent-exposed in the free protein. Similarly, Rpn13's strictly conserved F98 located on S6 also becomes less



**Figure 2 | Structure of Rpn13 Pru–ubiquitin complex defines a novel ubiquitin-binding motif.** **a**, Representative NOE interactions identified between Rpn13 Pru and ubiquitin. Each panel contains a selected region of a <sup>15</sup>N-dispersed nuclear Overhauser enhancement spectroscopy experiment recorded on <sup>15</sup>N-, <sup>13</sup>C- and 70% <sup>2</sup>H-labelled hRpn13 Pru mixed with equimolar quantities of unlabelled ubiquitin. All of the resonances displayed in this panel were unambiguously assigned as intermolecular NOE interactions with ubiquitin. Ubiquitin and Rpn13 Pru assignments are provided at the top and bottom of the expanded regions, respectively. **b**, Stereo representation of the mRpn13 Pru–ubiquitin complex oriented with ubiquitin at the top. At the interaction surface, secondary structural elements of ubiquitin and Rpn13 Pru are displayed in green and blue, respectively. Residues at the contact surface with intermolecular NOEs are in yellow (ubiquitin) or red (Rpn13 Pru), whereas those suggested to be at the

contact surface only by the NMR titration experiments are displayed in purple (ubiquitin) or cyan (Rpn13 Pru). M31, C88 and E111, which shift upon hRpn2 (amino acids 797–953) addition, are displayed in dark green. **c**, Specific amino-acid substitutions were made within the S2–S3, S4–S5 and S6–S7 loops of mRpn13 Pru by *in vitro* mutagenesis. The protein products were expressed as GST fusions and used in GST pull-down assays to highlight the importance of these loops for tetraubiquitin binding. **d**, Ubiquitin (blue), hRpn2 (amino acids 797–953) (green) or ubiquitin and hRpn2 (amino acids 797–953) (red) were added to the hRpn13 Pru domain, which was monitored by <sup>1</sup>H, <sup>15</sup>N heteronuclear single-quantum coherence experiments. Comparison with the spectrum acquired on the protein alone (black) indicates that S55, F76 and D78 bind ubiquitin in a manner that is independent of hRpn2 (amino acids 797–953), whereas M31, C88 and E111 bind hRpn2 in a ubiquitin-independent manner.

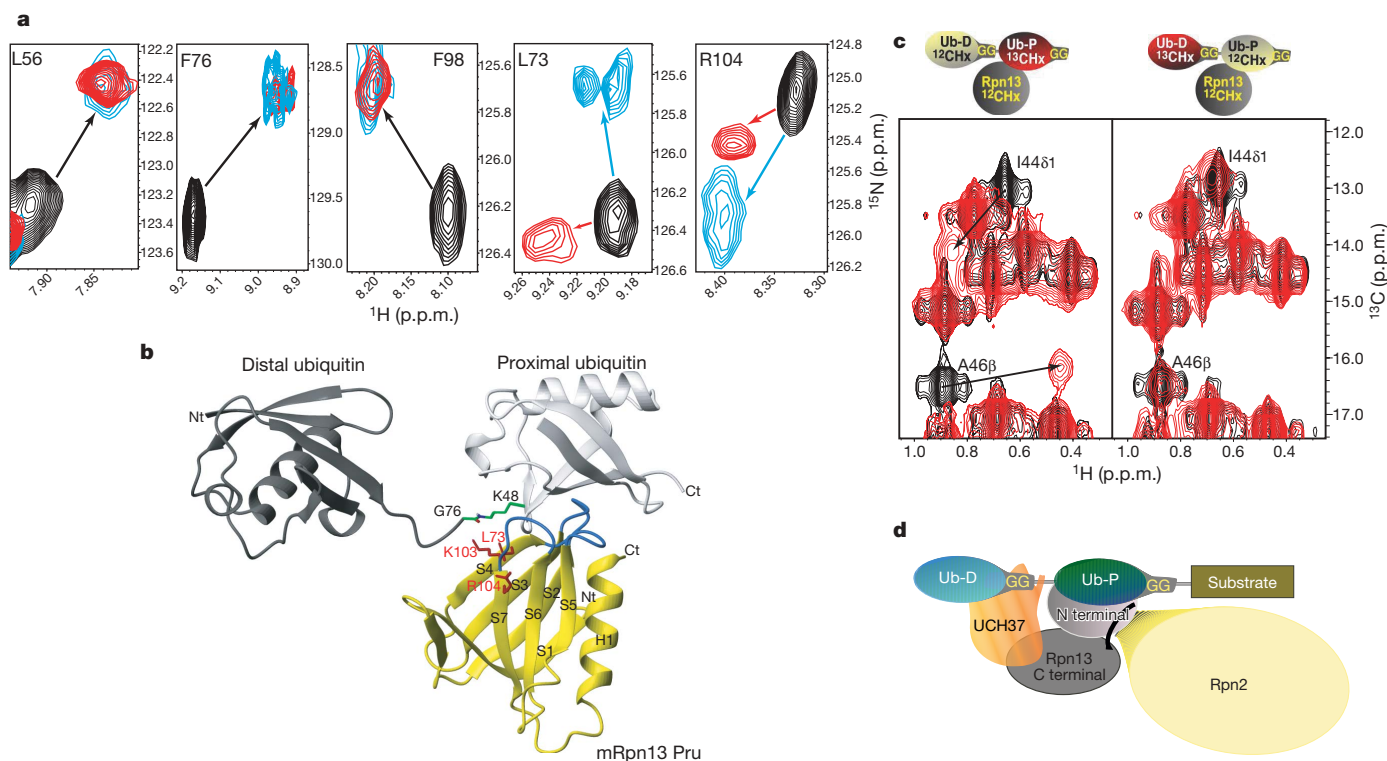
solvent-accessible through interactions with A46 and G47 of ubiquitin. Calculation of the electrostatic potential of mRpn13 illustrates that a hydrophobic region within its ubiquitin-binding surface that contains L56 and F76 is available to interact with ubiquitin's L8, I44 and V70 (Supplementary Fig. 3). Complementary electrostatic interactions between Rpn13 Pru and ubiquitin also stabilize the complex, including interactions of D78 and D79 of Rpn13 Pru with ubiquitin's H68, as well as D53 and D54 with ubiquitin's R42 and R72.

In total, the contact surface of Rpn13 Pru and ubiquitin comprises  $1256 \text{ \AA}^2$ , which is large for ubiquitin receptors. Thus, the relatively high affinity of hRpn13 binding to monoubiquitin<sup>8</sup> is partly explained by the enlarged contact surface compared with that of EAP45 GLUE ( $1000 \text{ \AA}^2$  in total)<sup>9</sup>. Published values for the total buried surface of Cue2-1<sup>cuc</sup> and Dsk2<sup>UBA</sup> upon ubiquitin binding are even smaller: 960 and  $800 \text{ \AA}^2$ , respectively<sup>10,11</sup>.

To analyse the significance of specific interactions identified in the mRpn13-ubiquitin complex, we made several amino-acid substitutions including L56A, I75R, F76R, D79N and F98R. The ubiquitin-binding competency of the resulting amino-acid-substituted proteins was tested by using GST-4xUb (created by the in-frame expression of glutathione S-transferase (GST) and four ubiquitin sequences) in pull-down assays (Fig. 2c). These experiments validated our mRpn13-ubiquitin structure and provided strong evidence for the importance of the S2-S3, S4-S5 and S6-S7 loops in

ubiquitin binding by mRpn13. In particular, the single amino-acid substitutions L56A, I75R, F76R and F98R abrogate ubiquitin binding, and a strong reduction is observed for the D79N mutation. We tested how three of these mutations affect mRpn13 structural integrity. In particular, NMR experiments were performed on mRpn13 Pru with the L56A, F76R or D79N mutations incorporated and compared with the wild-type mRpn13 Pru domain. These comparisons demonstrated that the loss of ubiquitin binding was not caused by loss of structural integrity (Supplementary Fig. 4). Altogether, our results demonstrate that ubiquitin binding is defined by key interactions with residues within the S2-S3, S4-S5 and S6-S7 loops.

To function as a proteasomal ubiquitin receptor, Rpn13 must bind ubiquitin and proteasome components simultaneously. In yeast and mammals, Rpn13 binds to Rpn2 through its Pru domain<sup>5,12,13</sup>. Although the Pru domain also binds ubiquitin, we found that Rpn2 binding does not disturb the Rpn13 loops that bind ubiquitin. By using a nested set of N-terminal deletions in human Rpn2 (hRpn2), we determined a fragment spanning amino acids 797-953 to bind mRpn13 (Supplementary Fig. 5). The addition of this fragment (hRpn2 (797-953)) to hRpn13 did not affect residues at the ubiquitin contact surface, which shift only upon ubiquitin addition, as demonstrated for S55, F76 and D78 (Fig. 2d). By contrast, M31, C88 and E111, which are unaffected by ubiquitin, shift after hRpn2 (797-953) addition. Furthermore, when both Rpn2 and ubiquitin were added, S55, F76 and D78 contact ubiquitin whereas M31, C88



**Figure 3 | Preferential binding to the proximal subunit of K48-linked diubiquitin by Rpn13 Pru allows Uch37 access to the distal subunit.**

**a**, Comparison of spectra acquired with monoubiquitin versus K48-linked tetraubiquitin reveals identical effects for Rpn13's L56, F76 and F98 but differences for L73 and R104. Monoubiquitin (red) or tetraubiquitin (cyan) was added to hRpn13, which was monitored by  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single-quantum coherence experiments. The spectrum of free hRpn13 is indicated in black, and the molar ratio of monoubiquitin (red) or tetraubiquitin (cyan) to hRpn13 was 1:1 in the represented spectra. **b**, Computer-generated model of the mRpn13 Pru-diubiquitin complex. White and grey ribbon diagrams display the proximal and distal ubiquitin, respectively, whereas a balls-and-sticks representation is used for the K48-G76 isopeptide bond linkage. Rpn13 Pru is coloured in yellow and loops recognizing ubiquitin in blue, whereas L73, K103 and R104 are displayed in red. Diubiquitin was created by

Insight II software based on atomic coordinates for the mRpn13 Pru-ubiquitin complex and monoubiquitin (PDB entry 1D3Z). In this model, the distal subunit of diubiquitin is positioned arbitrarily, as its only constraints prohibit steric clashes with other atoms. **c**, hRpn13 Pru interacts with the I44  $\delta_1$  and A46 methyl groups of the proximal subunit.  $^1\text{H}$ ,  $^{13}\text{C}$  heteronuclear multiple-quantum coherence spectra were acquired on samples containing either no (black) or equimolar unlabelled hRpn13 (amino acids 1-150) (red) mixed with diubiquitin in which its proximal (left) or distal (right) subunit is labelled with  $^{13}\text{C}$ . The shifted resonances are labelled. **d**, Model for how Rpn13 participates in Uch37 deconjugation of ubiquitinated substrates. Rpn13's C-terminal domain (grey) binds Uch37 (orange) as its N-terminal domain binds the polyubiquitin chain and Rpn2/S1 (yellow). In this model, Uch37 binds to the distal subunit (light blue) of the chain while Rpn13 binds the proximal subunit (dark blue).



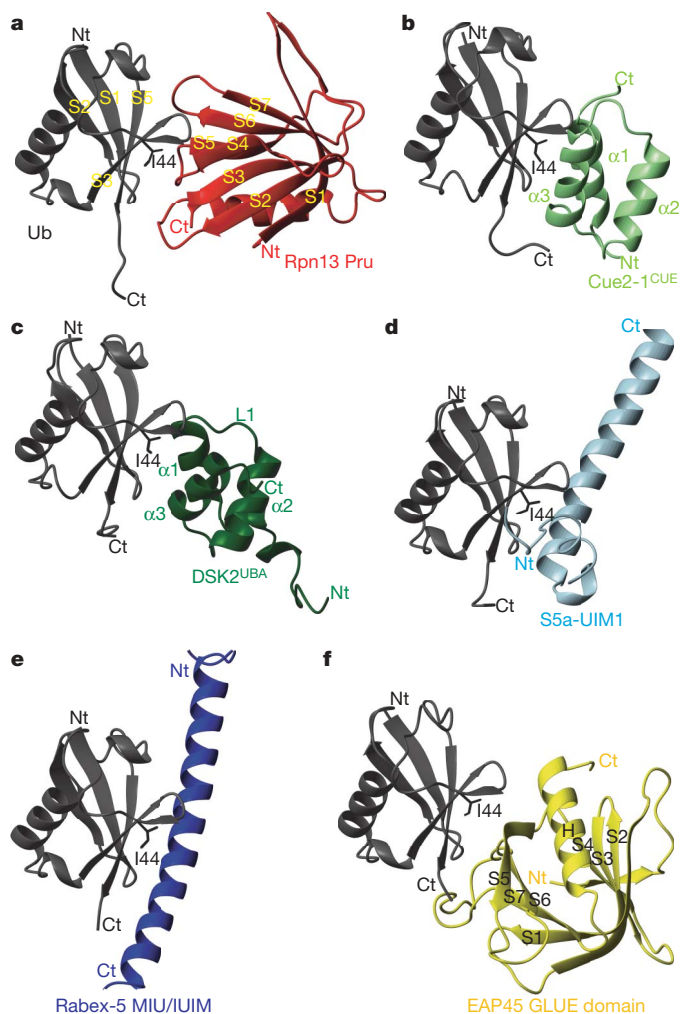
and E111 contact Rpn2 (Fig. 2d), indicating that the two binding surfaces are largely independent. M31, C88 and E111 are conserved in mRpn13 and map to S1, the S5–S6 loop, and the region linking S7 to H1, respectively. These elements are clustered in a region that is opposite to the ubiquitin-binding loops of Rpn13.

26S proteasomes exhibit high affinity for ubiquitinated substrates. Ubiquitin chains linked by isopeptide bonds between K48 and the C-terminal glycine of neighbouring ubiquitin molecules are known to trigger proteasomal degradation of the labelled protein<sup>14,15</sup>. We found that Rpn13 Pru binds K48-linked tetraubiquitin in a manner comparable to that of monoubiquitin. More specifically, tetraubiquitin and monoubiquitin caused chemical-shift changes to the same hRpn13 residues (Supplementary Fig. 6a), including L56, F76 and F98, and shifted them almost identically (Fig. 3a). Only two residues in hRpn13 exhibit changes that are specific to K48-linked tetraubiquitin, namely L73 and R104 (Fig. 3a and Supplementary Fig. 6a, red). These residues and the side-chain atoms of neighbouring K103 are proximal to each other; in the mRpn13–monoubiquitin structure they are directed towards the side-chain atoms of K48. This arrangement is congruent with Rpn13 binding the proximal subunit of

diubiquitin, namely that which forms an isopeptide using K48 (Fig. 3b). We tested this model further by monitoring the side-chain atoms of the proximal or distal subunit of diubiquitin upon addition of hRpn13 Pru. More specifically, unlabelled hRpn13 Pru was added to diubiquitin with either its proximal or distal subunit <sup>13</sup>C labelled, and the effect recorded by <sup>1</sup>H, <sup>13</sup>C heteronuclear multiple-quantum coherence experiments. Significant resonance shifting characteristic of hRpn13 binding was observed for the I44  $\gamma_1$ , I44  $\delta_1$  and A46 methyl groups of the proximal subunit (Fig. 3c, left panel, and Supplementary Fig. 7a). By contrast, resonances of the distal subunits exhibited only minor perturbations, most likely because of loss of intramolecular interactions with the proximal subunit. These data provide strong evidence that the major contacts formed between hRpn13 Pru and diubiquitin involve residues of the proximal subunit, at least when these two proteins are at equimolar concentration. Further evidence of this binding mode is provided by analytical ultracentrifugation data, which revealed 1:1 binding stoichiometry between hRpn13 Pru and diubiquitin (Supplementary Fig. 7b).

In conclusion, we reveal that the ubiquitin-binding region of proteasome subunit Rpn13 adopts a PHD fold, and we solve its structure complexed with ubiquitin to unveil a new ubiquitin-binding mode. PHDs are present in a remarkably large number of proteins<sup>16</sup>, but Rpn13 Pru is the first example of a PHD structure within the 26S proteasome. Rpn13, like many other ubiquitin receptors, binds to the L8, I44 and V70 hydrophobic pocket of ubiquitin (Fig. 2b). However, it is the first to bind this region using exclusively loops (Fig. 4a). Most of the ubiquitin receptors characterized so far use  $\alpha$ -helices to bind this surface of ubiquitin, including the ubiquitin-associated (UBA)-domain, coupling ubiquitin to endoplasmic reticulum degradation (CUE)-domain, ubiquitin-interacting motif (UIM), double-sided UIM (DUIM), inverted UIM (MIU), as well as GAT (GGA and TOM)-binding motifs (GGA, golgi-localized,  $\gamma$ -ear-containing, Arf (ADP-ribosylation factor)-binding protein; TOM, target of Myb). Among them, the UBA and CUE domains are structurally homologous, with a common three-helical bundle architecture. Cue2-1<sup>CUE</sup> binds ubiquitin through the  $\alpha 1$  and  $\alpha 3$  helices (Fig. 4b)<sup>10</sup>, whereas the Dsk2<sup>UBA</sup> uses the loop between  $\alpha 1$  and  $\alpha 2$ , as well as the C-terminal part of  $\alpha 3$  (Fig. 4c)<sup>11</sup>. Structural characterization of the UIMs demonstrated that a single  $\alpha$ -helix is sufficient for binding this region of ubiquitin<sup>17–19</sup>. The UIM helix includes a conserved alanine neighbored by a bulky hydrophobic residue, each of which packs against ubiquitin's I44 as demonstrated in the S5a UIM1–ubiquitin complex (Fig. 4d)<sup>19</sup>. Rabex-5 MIU/IUIM<sup>20,21</sup> and the polymerase  $\eta$  ubiquitin-binding zinc finger domain<sup>22</sup> similarly bind this region in ubiquitin through a single  $\alpha$ -helix, but in the reverse orientation (Fig. 4e). The GLUE domain of ESCRT-II EAP45, which exhibits a split pleckstrin-homology topology, is the only previously known ubiquitin-binding PHD<sup>23</sup>. However, it binds ubiquitin in a different manner: the I44-containing surface of ubiquitin is contacted by residues within secondary structural elements including the EAP45 C-terminal helix corresponding to H1 in Rpn13<sup>23</sup> (Fig. 4f). Moreover, although the longer S6–S7 loop of EAP45 is involved in binding ubiquitin, the S2–S3 and S4–S5 loops are not; instead, contacts are formed by residues from S5 and S6.

In addition to its unique monoubiquitin-binding mode, we have demonstrated that Rpn13 has a novel preference for diubiquitin elements within K48-linked chains<sup>8</sup> and that it most likely interacts directly with the isopeptide bond within a ubiquitin chain. This ubiquitin-binding mode is consistent with Rpn13's functional relationship with Uch37, which it adds to the collection of chain-processing enzymes in the proteasome's regulatory particle<sup>4–6</sup>. For diubiquitin, hRpn13 binding to the proximal subunit would leave the distal one available to interact with Uch37 (Fig. 3d). Evidence exists for Uch37 binding to the distal subunit of polyubiquitin; it is reported to be incapable of disassembling ubiquitin chains in which the distal ubiquitin contains the L8A and I44A mutations<sup>24</sup> and dismantles chains by removing one ubiquitin moiety at a time from the



**Figure 4 | Structural comparison of ubiquitin receptors complexed with ubiquitin.** **a–f**, Complex structures of ubiquitin and specific receptors displayed with ubiquitin in the same orientation (grey) and ubiquitin's I44 shown in black sticks. Each receptor has a unique colour coding: **a**, Rpn13 Pru (red); **b**, Cue2-1<sup>CUE</sup> (PDB code 1OTR; light green); **c**, Dsk2<sup>UBA</sup> (PDB code 1WR1; dark green); **d**, S5a-UIM1 (PDB code 1YX5; light blue); **e**, Rabex-5 MIU/IUIM (PDB code 2FIF; dark blue); **f**, EAP45 GLUE domain (PDB code 2HTH; yellow). All structures are compared by a best-fit superposition of bound ubiquitin (grey). In **c**, L1 denotes the loop connecting  $\alpha 1$  and  $\alpha 2$ .

distal end<sup>25</sup>. Uch37's distal-end deconjugation of ubiquitin chains<sup>25</sup> complements that of Ubp6 and Rpn11, as Ubp6 can deconjugate multiple ubiquitins in a single cleavage event<sup>26</sup> and Rpn11 performs 'en bloc' deubiquitination from the proximal end<sup>27,28</sup>. Deubiquitinating activities, particularly that of Ubp6, are antagonized by another regulatory particle component, the chain elongation factor Hul5<sup>29</sup>. With so many receptors and chain-processing enzymes within the regulatory particle, the detailed pathway by which a substrate is degraded may be subject to many stochastic variations. Whether this unanticipated design promotes high substrate flux through the proteasome is unclear, but it seems well suited to allow the cell to fine-tune proteasome activity.

## METHODS SUMMARY

**Expression, purification, crystallization and structure determination of mRpn13 Pru.** mRpn13 Pru was overexpressed in *Escherichia coli* strain BL21(DE3) RIL (Stratagene) and purified by GST-affinity chromatography using a PreScission Protease (GE Healthcare) cleavage site followed by size-exclusion chromatography. Rpn13 Pru was crystallized by the hanging-drop vapour-diffusion method and frozen in a stream of liquid nitrogen during X-ray exposure. Single anomalous dispersion methods were performed using synchrotron radiation at the BW6 beamline at the DESY-centre in Hamburg, Germany. Native data were collected to 1.7-Å resolution (Supplementary Table 1). Details about recombinant DNA modifications, expression and purification of mutant forms of Rpn13 Pru, data processing, phase determination, model building and structural refinement are described in Supplementary Information. **NMR spectroscopy.** Chemical shift assignments and spectra were acquired as described in Supplementary Information. mRpn13 Pru-ubiquitin and mRpn13 Pru-diubiquitin complexes were generated as described in Supplementary Information.

Received 31 October 2007; accepted 18 March 2008.

- Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059–7061 (1994).
- Fu, H. *et al.* Multiubiquitin chain binding and protein degradation are mediated by distinct domains within the 26 S proteasome subunit Mcb1. *J. Biol. Chem.* **273**, 1970–1981 (1998).
- Jorgensen, J. P. *et al.* Adrm1, a putative cell adhesion regulating protein, is a novel proteasome-associated factor. *J. Mol. Biol.* **360**, 1043–1052 (2006).
- Yao, T. *et al.* Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nature Cell Biol.* **8**, 994–1002 (2006).
- Hamazaki, J. *et al.* A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J.* **25**, 4524–4536 (2006).
- Qiu, X. B. *et al.* hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *EMBO J.* **25**, 5742–5753 (2006).
- Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
- Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* doi:10.1038/nature06926 (this issue) (2008).
- Hirano, S. *et al.* Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain. *Nature Struct. Mol. Biol.* **13**, 1031–1032 (2006).
- Kang, R. S. *et al.* Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell* **113**, 621–630 (2003).
- Ohno, A. *et al.* Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. *Structure* **13**, 521–532 (2005).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).

- Gandhi, T. K. *et al.* Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nature Genet.* **38**, 285–293 (2006).
- Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583 (1989).
- Finley, D. *et al.* Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant. *Mol. Cell. Biol.* **14**, 5501–5509 (1994).
- Lemmon, M. A. Pleckstrin homology domains: not just for phosphoinositides. *Biochem. Soc. Trans.* **32**, 707–711 (2004).
- Fisher, R. D. *et al.* Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.* **278**, 28976–28984 (2003).
- Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L. & Radhakrishnan, I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J.* **22**, 4597–4606 (2003).
- Wang, Q., Young, P. & Walters, K. J. Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J. Mol. Biol.* **348**, 727–739 (2005).
- Lee, S. *et al.* Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *Nature Struct. Mol. Biol.* **13**, 264–271 (2006).
- Penengo, L. *et al.* Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell* **124**, 1183–1195 (2006).
- Bomar, M. G., Pai, M. T., Tzeng, S. R., Li, S. S. & Zhou, P. Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase  $\eta$ . *EMBO Rep.* **8**, 247–251 (2007).
- Alam, S. L. *et al.* Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nature Struct. Mol. Biol.* **13**, 1029–1030 (2006).
- Lam, Y. A., DeMartino, G. N., Pickart, C. M. & Cohen, R. E. Specificity of the ubiquitin isopeptidase in the PA700 regulatory complex of 26 S proteasomes. *J. Biol. Chem.* **272**, 28438–28446 (1997).
- Lam, Y. A., Xu, W., DeMartino, G. N. & Cohen, R. E. Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature* **385**, 737–740 (1997).
- Hanna, J. *et al.* Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127**, 99–111 (2006).
- Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–615 (2002).
- Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
- Crosas, B. *et al.* Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. *Cell* **127**, 1401–1413 (2006).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The help of G. Bourenkov (DESY, BW6, Hamburg, Germany) during synchrotron data collection is gratefully acknowledged. We also thank J. Lary, J. Cole and the National Analytical Ultracentrifugation Facility of the University of Connecticut for performing the sedimentation experiments. NMR data were acquired at the University of Minnesota and the data processed in the Minnesota Supercomputing Institute's Basic Sciences Computing Laboratory. This work was supported by National Institutes of Health CA097004 (K.W.), GM43601 (D.F.) and GM008700-Chemistry Biology Interface Training Grant (L.R.), Deutsche Forschungsgemeinschaft (DI 931/3-1) and the Cluster of Excellence 'Macromolecular Complexes' of the Goethe University Frankfurt (EXC115) to I.D., and Deutsche Forschungsgemeinschaft SFB740/TP B4 (M.G.).

**Author Contributions** The crystal structure of mRpn13 was solved by P.S. and M.G. (Fig. 1); the complexed structure of mRpn13-ubiquitin (Figs 2a, b and 4), the model structure of mRpn13-diubiquitin (Fig. 3), and NMR studies (Fig. 2d) were conducted by X.C., L.R., N.Z. and K.J.W.; and identification of the interactions, minimal binding domains, protein purification and western blot analysis (Fig. 2c) were performed by K.H. Plasmids were created by K.H. and S.E. I.D., M.G., D.F. and K.J.W. all contributed to design of the experiments and writing the manuscript.

**Author Information** Coordinates and structure factors of mRpn13 Pru and mRpn13 Pru-ubiquitin have been deposited in Protein Data Bank under accession numbers 2R2Y and 2Z59, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.G. ([michael.groll@ch.tum.de](mailto:michael.groll@ch.tum.de)), K.J.W. ([walte048@umn.edu](mailto:walte048@umn.edu)) or I.D. ([ivan.dikic@biochem2.de](mailto:ivan.dikic@biochem2.de)).

# A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*

Anastasios D. Tsaousis<sup>1,2</sup>, Edmund R. S. Kunji<sup>3</sup>, Alina V. Goldberg<sup>1</sup>, John M. Lucocq<sup>4</sup>, Robert P. Hirt<sup>1</sup> & T. Martin Embley<sup>1</sup>

Mitochondria use transport proteins of the eukaryotic mitochondrial carrier family (MCF) to mediate the exchange of diverse substrates, including ATP, with the host cell cytosol. According to classical endosymbiosis theory, insertion of a host-nuclear-encoded MCF transporter into the protomitochondrion was the key step that allowed the host cell to harvest ATP from the enslaved endosymbiont<sup>1</sup>. Notably the genome of the microsporidian *Encephalitozoon cuniculi* has lost all of its genes for MCF proteins<sup>2</sup>. This raises the question of how the recently discovered microsporidian remnant mitochondrion, called a mitosome, acquires ATP to support protein import and other predicted ATP-dependent activities<sup>2–4</sup>. The *E. cuniculi* genome does contain four genes for an unrelated type of nucleotide transporter used by plastids and bacterial intracellular parasites, such as *Rickettsia* and *Chlamydia*, to import ATP from the cytosol of their eukaryotic host cells<sup>5–7</sup>. The inference is that *E. cuniculi* also uses these proteins to steal ATP from its eukaryotic host to sustain its lifestyle as an obligate intracellular parasite. Here we show that, consistent with this hypothesis, all four *E. cuniculi* transporters can transport ATP, and three of them are expressed on the surface of the parasite when it is living inside host cells. The fourth transporter co-localizes with mitochondrial Hsp70 to the *E. cuniculi* mitosome. Thus, uniquely among eukaryotes, the traditional relationship between mitochondrion and host has been subverted in *E. cuniculi*, by reductive evolution and analogous gene replacement. Instead of the mitosome providing the parasite cytosol with ATP, the parasite cytosol now seems to provide ATP for the organelle.

Microsporidia have undergone extreme genomic and cellular reduction as obligate intracellular parasites of other eukaryotes<sup>2,8</sup>. The published genome<sup>2</sup> of *E. cuniculi* reveals that genes required for key energy-generating reactions, including the tricarboxylic acid cycle, fatty acid beta oxidation, the respiratory electron transport chain and the ATP synthase complex, are all absent. Thus, ATP production in *E. cuniculi* is only possible by substrate level phosphorylation<sup>2</sup>. Because proliferating microsporidia recruit host mitochondria near their plasma membrane, it has been proposed that these parasites could use host-derived ATP to supplement their energy budget<sup>9</sup>. The presence in the genome of *E. cuniculi* of genes encoding homologues of ADP/ATP transporters used by bacterial parasites to steal ATP provides a potential means of achieving this goal<sup>2</sup>.

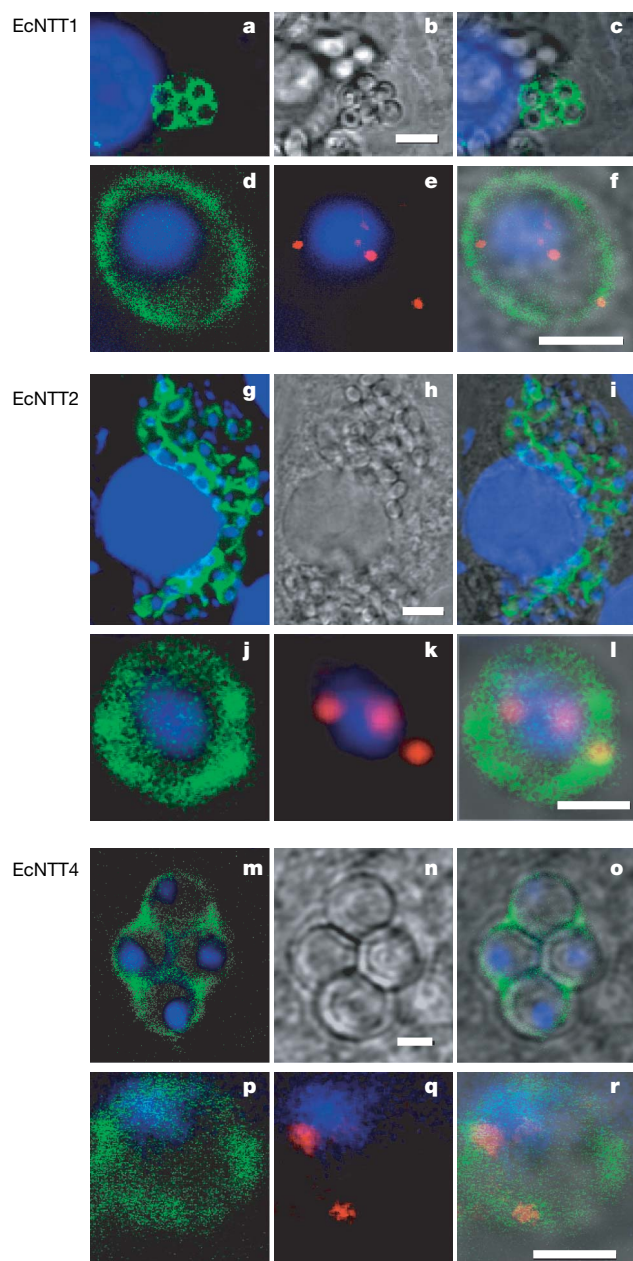
We compared the sequences of the four *E. cuniculi* putative transporters<sup>2</sup> to those of previously characterized ATP/ADP translocases or nucleotide transporters (NTTs) from plastids and bacterial parasites<sup>5,6,10,11</sup>. The four *E. cuniculi* sequences (EcNTT1 to EcNTT4) are divergent from each other (28–36% identity) and are predicted to

contain 11–12 transmembrane domains, as commonly observed for bacterial and plastid homologues<sup>5</sup> (Supplementary Fig. 1). There was no *in silico* evidence for any compartment-specific targeting signals (Supplementary Fig. 1). To investigate the expression and cellular location of the *E. cuniculi* transporters, we raised antisera against variable regions of each protein (Supplementary Fig. 2). All four antisera detected specific bands in the protein extracts from infected rabbit kidney cells, but no bands were detected in protein extracts from uninfected cells (Supplementary Fig. 3). The antiserum to EcNTT2 also detected a strong band in protein extracts from spores, indicating that EcNTT2 is present at this stage of the parasite lifecycle. The other transporters gave faint or no signal for spore protein extracts (Supplementary Fig. 3).

To determine the cellular location of the *E. cuniculi* transporters, indirect immunofluorescence analyses (IFAs) were carried out on infected rabbit kidney cells and on isolated parasites. EcNTT1, EcNTT2 and EcNTT4 gave a labelling distribution that was consistent with them being present on the surface of parasites living inside host cells (Fig. 1), but these transporters could not be detected in purified parasites (data not shown). In contrast, EcNTT3 was localized to discreet areas of the cytosol of *E. cuniculi* living inside host cells, and produced the same labelling for most (~70%, data not shown) purified parasites (Fig. 2). These data suggested that EcNTT3 might be targeted to the mitosomes of *E. cuniculi*. To test this hypothesis, we generated a specific antiserum to *E. cuniculi* mitochondrial heat-shock protein 70 (mtHsp70). This protein is a classic mitochondrial marker<sup>12</sup> and the homologous protein is already known to locate exclusively to mitosomes of the microsporidian *Trachipleistophora hominis*<sup>3</sup>. In reciprocal labelling experiments with *T. hominis*, specific antisera for either the *E. cuniculi* or the *T. hominis* mtHsp70 produced the same pattern of labelling for both species (Supplementary Fig. 4), validating mtHsp70 as a marker for *E. cuniculi* mitosomes. Consistent with the IFA results, transmission electron microscopy of *E. cuniculi* cells identified double-membrane-bounded organelles with morphology and size (66 nm by 110 nm) similar to *T. hominis* mitosomes<sup>3</sup> (Supplementary Fig. 5). The mitosomes were often close to structures resembling spindle pole bodies<sup>13</sup>, which are involved in cell division<sup>14</sup>; this close physical juxtaposition may aid segregation of mitosomes during cell division. There was no evidence of any overlap between the signals for EcmtHsp70 and EcNTT1, EcNTT2 or EcNTT4 (Fig. 1). However, the fluorescence signals associated with EcmtHsp70 and EcNTT3 clearly overlapped (Fig. 2), indicating that EcNTT3 is localized exclusively to *E. cuniculi* mitosomes. The structural features of the mitochondrial EcNTT3 protein suggest that, like MCF proteins of mitochondria<sup>15</sup>, it is a multi-spanning membrane protein. However, so far we have

<sup>1</sup>Institute for Cell and Molecular Biosciences, Catherine Cookson Building, Framlington Place, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. <sup>2</sup>School of Biology, Devonshire Building, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. <sup>3</sup>The Dunn Human Nutrition Unit, Medical Research Council, Cambridge CB2 2XY, UK. <sup>4</sup>School of Life Sciences, WTB/MSI complex, University of Dundee, Dundee DD1 5EH, UK.



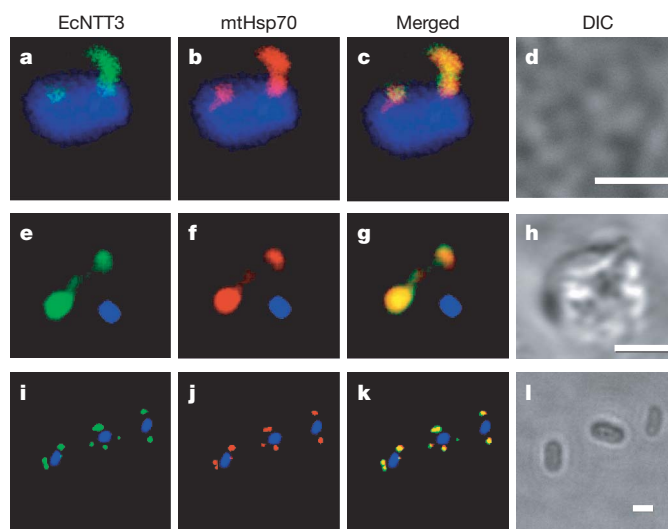


**Figure 1 | Cellular localization of *E. cuniculi* transporters *in situ* using immunofluorescence and confocal microscopy.** **a**, Six *E. cuniculi* cells inside a rabbit kidney host cell are labelled with polyclonal antisera to EcNTT1 (green), consistent with a location of EcNTT1 at the plasma membrane of the parasite. The host nucleus is labelled with DAPI (4,6-diamidino-2-phenylindole, blue). **b**, A differential interference contrast (DIC) image of the same field (scale bar, 4  $\mu$ m). **c**, Merge of **a** and **b**. **d**, A single parasite labelled with antisera to EcNTT1 (green), in which the parasite nucleus is stained blue with DAPI. **e**, The same cell labelled with antisera to *E. cuniculi* mtHsp70 (red) detects three mitochondria. **f**, Merge of **d** and **e** superimposed over a DIC image shows no co-localization of the two labels (scale bar, 1.49  $\mu$ m). **g**, *E. cuniculi* cells at different stages of the parasite lifecycle includes cells labelled with antisera to EcNTT2 (green), consistent with a location of EcNTT2 at the plasma membrane of the parasite. **h**, DIC image of the same cells. **i**, Merge of **g** and **h** (scale bar, 4  $\mu$ m). **j**, A single parasite labelled with antisera to EcNTT2. **k**, The same cell labelled with the polyclonal antisera to *E. cuniculi* mtHsp70 (red). **l**, Merge of **j** and **k**, showing that the two antisera label different structures (scale bar, 1.49  $\mu$ m). **m**, Four *E. cuniculi* cells inside a rabbit kidney cell labelled with antisera to EcNTT4 (green), consistent with a location at the plasma membrane of the parasite. **n**, DIC image of the same field (scale bar, 2  $\mu$ m). **o**, Merge of **m** and **n**. **p**, A single parasite labelled with polyclonal antisera to EcNTT4. **q**, The antisera to Hsp70 labels two mitochondria. **r**, Merge **p** and **q** (scale bar, 1.49  $\mu$ m).

been unable to develop efficient immuno-electron microscopy localization techniques for *E. cuniculi* to investigate this prediction at the ultrastructural level.

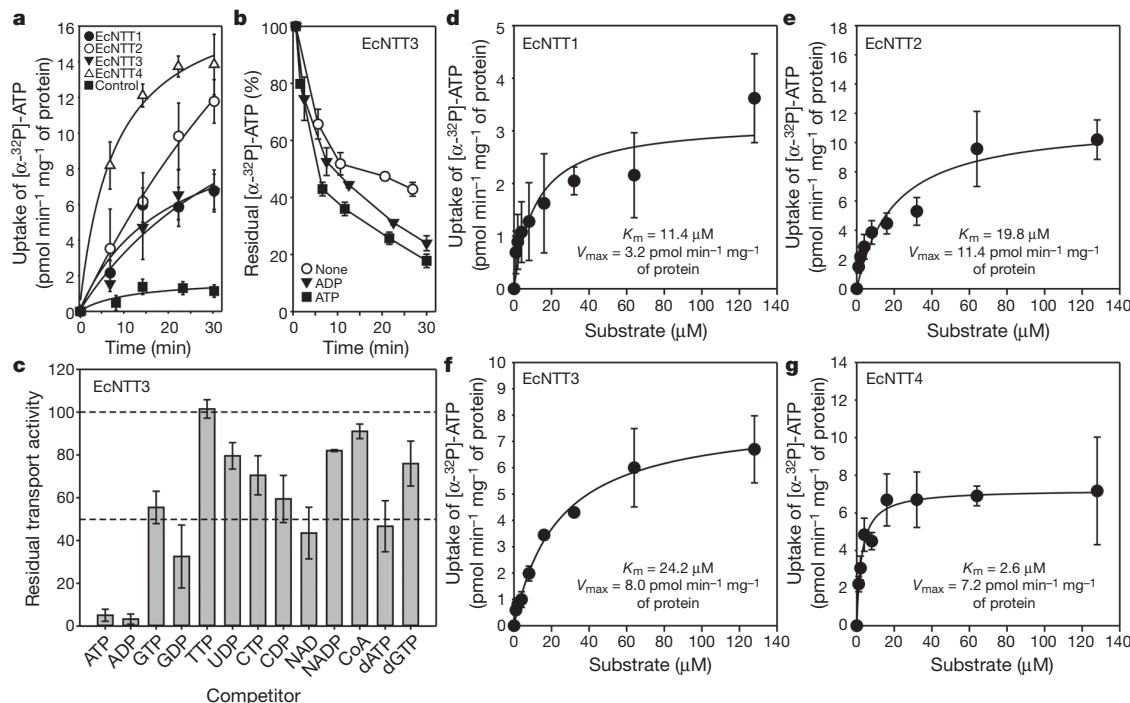
The *E. cuniculi* proteins contain the charged residues conserved among NTTs that are crucial for the functioning of the *Arabidopsis* plastid ATP/ADP transporters<sup>10</sup> (Supplementary Fig. 1). There are no homologous genetic systems for characterizing microsporidian proteins so we used the *Escherichia coli* expression system, previously used to characterize *Chlamydia*, *Rickettsia* and plastid NTTs<sup>6,7,11,16</sup>, to obtain functional data for the *E. cuniculi* transporters. All four *E. cuniculi* transporters mediated uptake of radiolabelled ATP above background levels when expressed in *E. coli*, with EcNTT4 mediating the fastest import (Fig. 3a). The dose-response curves (Fig. 3d–g) showing the effect of increasing substrate concentration on the uptake of ATP, indicated an apparent Michaelis constant ( $K_m$ ) of 11.4  $\mu$ M, 19.8  $\mu$ M, 24.2  $\mu$ M and 2.6  $\mu$ M for EcNTT1, EcNTT2, EcNTT3 and EcNTT4, respectively. These data show that all four EcNTTs have a high affinity for ATP. Analysis of the substrate specificity of EcNTT3 (Fig. 3c) and of EcNTT1, EcNTT2 and EcNTT4 (Supplementary Fig. 6) showed that the uptake of radiolabelled ATP was substantially reduced by competition with excess ATP or ADP and to a much lesser extent by competition with other nucleotides. These data demonstrate that all four EcNTTs are able to transport ADP and ATP, in contrast to the situation in bacterial parasites in which typically only one paralogue will mediate ATP uptake<sup>7,11</sup>. Adenine nucleotide efflux was observed for EcNTT3 when external substrate was removed, indicating that this transporter is able to equilibrate nucleotide pools across a concentration gradient (Fig. 3b). Efflux was stimulated by the addition of external cold ATP or ADP, confirming that EcNTT3 is an exchanger of adenine nucleotides (Fig. 3b).

The observation that EcNTT1, EcNTT2 and EcNTT4 can transport ATP may help to explain why these transporters were not detected on the surface of purified parasites using IFA, but were only detected on the surface of parasites living inside host cells (Fig. 1). Reduced expression of surface-membrane-associated NTTs in



**Figure 2 | Evidence that *E. cuniculi* EcNTT3 is targeted to mitochondria.**

**a**, **e**, The polyclonal antisera to EcNTT3 (green) labels discrete structures within the cytosol of two different *E. cuniculi* cells, one per field. The nucleus of each parasite is labelled blue with DAPI. **b**, **f**, The polyclonal antisera to *E. cuniculi* mtHsp70 (red) identifies the structures as mitochondria. **c**, **g**, Merged images showing the overlap of both signals. **d**, **h**, The corresponding DIC image. **i**, The polyclonal antibody to EcNTT3 (green) labels discrete structures within the cytosol of three *E. cuniculi* purified cells. **j**, The polyclonal antibody to *E. cuniculi* mtHsp70 (red) identifies the structures as mitochondria. **k**, A merged image showing the overlap of signals. **l**, The corresponding DIC image. Scale bars, 1.49  $\mu$ m.



**Figure 3 | Transport of radiolabelled ATP by *E. coli* cells expressing *E. cuniculi* transporters.** **a**, Time dependency of  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  uptake into IPTG-induced *E. coli* cells harbouring pET16b with gene insert encoding EcNTT1, EcNTT2, EcNTT3 or EcNTT4, or control without insert. Induced cells were incubated with 2 nM  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  in the presence of 1  $\mu\text{M}$  cold ATP for the indicated time periods. Termination of uptake was carried out by quenching with ice-cold  $\text{KPi}$  buffer, followed immediately by rapid filtration. **b**, Back-exchange properties of EcNTT3 expressed in *E. coli* cells. Cells were induced and pre-loaded with 1  $\mu\text{M}$   $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  for 20 min, and were washed and diluted in  $\text{KPi}$  buffer with no substrate or with 250  $\mu\text{M}$  ADP or ATP. At specified time intervals, efflux was terminated as in **a**. **c**, The effect of different substrates on uptake of  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  by *E. coli* expressing

EcNTT3 was measured at a substrate concentration of 2 nM  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$ . Unlabelled effectors were present at a 50,000-fold excess over the labelled substrate. Termination of the experiment was carried out as in **a**. Rates of nucleotide uptake are given as a percentage of rates in the absence of excess nucleotide. **d–g**, Substrate saturation of  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  uptake by cells expressing EcNTT1–EcNTT4. IPTG-induced *E. coli* cells expressing EcNTT1 (**d**), EcNTT2 (**e**), EcNTT3 (**f**) or EcNTT4 (**g**) were incubated for 15 min with 2 nM  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$  and increasing concentrations of cold ATP, and were processed as described in Methods. The  $V_{\text{max}}$  and apparent  $K_m$  were calculated as described in Methods. Data points and errors bars for all panels represent the mean and s.d. of three independent experiments, respectively.

response to a reduced ATP status of the host cell cytosol, or of the surrounding milieu, has already been described for plastids and intracellular bacteria<sup>5,6</sup>. Shutdown is thought to occur to prevent loss of bacterial or plastid ATP by NTT-mediated transport down a concentration gradient. The purified *E. cuniculi* cells are mainly emerging spore stages<sup>17</sup> that are differentiating into dormant cells to survive exposure to the external environment. Thus, it would also make sense if *E. cuniculi* reduced expression of its cell-surface-associated NTTs as part of this differentiation process.

The use of bacterial-like nucleotide transporters to acquire ATP from another eukaryotic cell is a unique strategy for a eukaryotic parasite. Other intracellular parasites, such as *Leishmania* and *Plasmodium*, use transporters that are homologous to host proteins to cater for their energy and nucleotide needs<sup>15,18</sup>. A gene for an NTT has been reported for the microsporidian *Paranosema grylli* (GenBank accession number CAI30461), and preliminary data using the heterologous antisera to the *E. cuniculi* transporters suggest that NTTs are present on the surface of *T. hominis* (Supplementary Fig. 7). These species represent three of the five major microsporidian groups<sup>19</sup>, suggesting that the use of NTTs may be a general strategy used by microsporidians to steal ATP from their hosts, and that acquisition of NTTs was thus a key innovation supporting their lifestyle as obligate intracellular energy parasites. Consistent with this idea, phylogenetic analyses of the available microsporidian sequences suggest that NTT were acquired only once by microsporidia early in their evolution (Supplementary Fig. 8). The very patchy distribution of NTTs in diverse prokaryotes and eukaryotes does, however, suggest that lateral gene transfer has been involved in the evolution of intracellular energy parasitism<sup>20</sup>. Because there are no homologues of

NTTs in humans or other animals, these proteins may also represent promising drug targets<sup>21</sup>. New highly specific drugs are of interest because microsporidia such as *Enterocytozoon* remain a significant health problem for immunodeficient patients, including those with AIDS<sup>22,23</sup>.

The mitochondria of microsporidia are tiny and structurally undifferentiated<sup>3</sup>, they lack a genome and their protein import pathway is rudimentary<sup>2,24</sup>. On the basis of the published genome of *E. cuniculi*<sup>2</sup>, there are no obvious ways for the organelle to make or to import ATP. However, the mitochondria of *E. cuniculi* and *T. hominis* contain mitochondrial Hsp70, which requires ATP for its activity<sup>12</sup>. Our data suggest that *E. cuniculi* has evolved a unique solution to meeting its cellular and mitochondrial energy requirements, by using NTT proteins to exploit the ATP pool of its eukaryotic host, and of its own cytosol.

## METHODS SUMMARY

*E. cuniculi* strain II was grown in cultured rabbit kidney cells<sup>25</sup>. Guided by the published genome of *E. cuniculi*<sup>2</sup>, we PCR-amplified and cloned full-length coding sequences of four putative transporters from genomic DNA of *E. cuniculi* strain II. Antisera to cloned constructs comprising the most variable regions of each *E. cuniculi* transporter, and avoiding predicted transmembrane domains, were made in rats using standard methods. Antiserum to the full-length *E. cuniculi* mtHsp70 protein was prepared in rabbit. The specificity of each antiserum was confirmed by western blots of total proteins extracted from spores of *E. cuniculi*, rabbit kidney cells infected with *E. cuniculi*, and non-infected rabbit kidney cells. To determine the cellular location of the *E. cuniculi* transporters and *E. cuniculi* mtHsp70, we used IFA. Mammalian cells infected with *E. cuniculi* were grown and fixed on coverslips. Cells of *E. cuniculi* were also purified by disruption of infected rabbit kidney cells followed by filtration through glass wool and fixation on glass slides. After blocking, the attached cells were

incubated with the relevant antisera raised for each *E. cuniculi* protein and subsequently incubated with the fluorescent-dye-conjugated (Alexa red 594 or Alexa green 488) secondary antibodies. The labelled cells were visualized under a laser scanning confocal microscope. To obtain functional data for the *E. cuniculi* transporters, we used the *E. coli* expression system, previously used to characterize the homologous transport proteins from bacteria and plastids<sup>6,7,11,16</sup>. Dose-response curves for uptake of radiolabelled ATP by the four *E. cuniculi* transporters, competition experiments to investigate substrate specificity, and back-exchange assays for EcNTT3 using cold ADP or ATP were carried out by modified published methods<sup>6,7,11,16</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 February; accepted 7 March 2008.

Published online 30 April 2008.

- John, P. & Whatley, F. R. *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254**, 495–498 (1975).
- Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
- Williams, B. A., Hirt, R. P., Lucocq, J. M. & Embley, T. M. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* **418**, 865–869 (2002).
- Goldberg, A. V. *et al.* Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. *Nature* **452**, 624–628 (2008).
- Winkler, H. H. & Neuhaus, H. E. Non-mitochondrial ATP transport. *Trends Biochem. Sci.* **24**, 64–68 (1999).
- Haferkamp, I. *et al.* A candidate NAD<sup>+</sup> transporter in an intracellular bacterial symbiont related to *Chlamydiae*. *Nature* **432**, 622–625 (2004).
- Audia, J. P. & Winkler, H. H. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J. Bacteriol.* **188**, 6261–6268 (2006).
- Keeling, P. J. & Fast, N. M. Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu. Rev. Microbiol.* **56**, 93–116 (2002).
- Weidner, E., Findley, A. M., Dolgikh, V. & Skolova, J. in *The Microsporidia and Microsporidiosis* (eds Wittner, M. & Weiss, L. M.) (American Society of Microbiology, Washington DC, 1999).
- Trentmann, O., Decker, C., Winkler, H. H. & Neuhaus, H. E. Charged amino-acid residues in transmembrane domains of the plastidic ATP/ADP transporter from *Arabidopsis* are important for transport efficiency, substrate specificity, and counter exchange properties. *Eur. J. Biochem.* **267**, 4098–4105 (2000).
- Haferkamp, I. *et al.* Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol. Microbiol.* **60**, 1534–1545 (2006).
- Pfanner, N. & Geissler, A. Versatility of the mitochondrial protein import machinery. *Nature Rev. Mol. Cell Biol.* **2**, 339–349 (2001).
- Vavra, J. “Polar vesicles” of microsporidia are mitochondrial remnants (“mitosomes”)? *Folia Parasitol. (Praha)* **52**, 193–195 (2005).
- Zheng, L., Schwartz, C., Magidson, V., Khodjakov, A. & Olfierenko, S. The spindle pole bodies facilitate nuclear envelope division during closed mitosis in fission yeast. *PLoS Biol.* **5**, e170 (2007).
- Kunji, E. R. The role and structure of mitochondrial carriers. *FEBS Lett.* **564**, 239–244 (2004).
- Mohlmann, T. *et al.* Occurrence of two plastidic ATP/ADP transporters in *Arabidopsis thaliana* L.: molecular characterisation and comparative structural analysis of similar ATP/ADP translocators from plastids and *Rickettsia prowazekii*. *Eur. J. Biochem.* **252**, 353–359 (1998).
- Taupin, V., Metenier, G., Vivares, C. P. & Prensier, G. An improved procedure for Percoll gradient separation of sporogonial stages in *Encephalitozoon cuniculi* (Microsporidia). *Parasitol. Res.* **99**, 708–714 (2006).
- de Koning, H. P., Bridges, D. J. & Burchmore, R. J. S. Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy. *FEMS Microbiol. Rev.* **29**, 987–1020 (2005).
- Vossbrinck, C. R. & Debrunner-Vossbrinck, B. A. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol. (Praha)* **52**, 131–142 (2005).
- Richards, T. A., Hirt, R. P., Williams, B. A. & Embley, T. M. Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17–32 (2003).
- Schmitz-Esser, S. *et al.* ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J. Bacteriol.* **186**, 683–691 (2004).
- Didier, E. S. & Weiss, L. M. Microsporidiosis: current status. *Curr. Opin. Infect. Dis.* **19**, 485–492 (2006).
- Farthing, M. J. Treatment options for the eradication of intestinal protozoa. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **3**, 436–445 (2006).
- Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
- Hollister, W. S., Canning, E. U. & Anderson, C. L. Identification of Microsporidia causing human disease. *J. Eukaryot. Microbiol.* **43**, 104S–105S (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Booth for help with confocal microscopy, J. James for help with electron microscopy, I. Haferkamp for providing details about the transport assay for NTTs expressed in *E. coli* and S. Harris for advice on bioinformatics analyses. A.D.T. was supported by a scholarship from the Division of Biology, Newcastle University. T.M.E. acknowledges support from the British Royal Society and the Leverhulme Trust. E.R.S.K. is supported by the Medical Research Council UK.

**Author Contributions** A.D.T. cloned the EcNTT1–4 and mtHsp70 genes, generated the antisera and performed the western blots and IFAs; E.R.S.K. and A.D.T. performed the transport assays and their analysis; A.V.G. supervised and contributed to the cloning and antisera generation, western blots and IFAs; J.M.L. performed the electron microscopy work; and R.P.H. and T.M.E. designed the project, supervised the project on a day-to-day basis, and wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** The EcNTT1–4 and EcmtHsp70 sequences have been deposited in GenBank under accession numbers EU040266–EU040270. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.P.H. ([r.p.hirt@ncl.ac.uk](mailto:r.p.hirt@ncl.ac.uk)) or E.R.S.K. ([ek@mrc-dunn.cam.ac.uk](mailto:ek@mrc-dunn.cam.ac.uk)).



## METHODS

**Growth of *E. cuniculi* in rabbit kidney cells.** Rabbit kidney cells (RK-13) were infected with *E. cuniculi* strain II (provided by E. S. Didier) and grown as described<sup>25</sup>.

**Generation of antisera to *E. cuniculi* transporters EcNTT1–4 and EcmtHsp70.** The predicted amino acid sequences of the four *E. cuniculi* strain II transporters were identical to those predicted for *E. cuniculi* GB-M1. Three variable regions of sequence located outside of predicted transmembrane domains (Supplementary Fig. 1) were identified for each EcNTT (Supplementary Fig. 2). PCR was used to amplify each set of three fragments, with appropriate restriction digestion sites at their termini. After amplification and restriction digestion, each set of fragments was ligated to encode four different fusion peptides, one for each transporter. The fusion peptides exhibited between 13% and 31% identity to each other. After sequencing, each construct was cloned into the pQE-40 expression vector (Qiagen) and expressed in *E. coli* M15 (pREP4) cells (Qiagen). Each histidine-tagged recombinant protein was extracted from inclusion bodies using Bugbuster reagent (Novagen) and was purified by gel electrophoresis. One milligram of each protein was used for the commercial production of rat polyclonal antisera (Eurogentec).

The entire *E. cuniculi* strain II mtHsp70 open reading frame was PCR-amplified, cloned into pET30a (Novagen), and sequenced and transformed into *E. coli* C43 (DE3) cells<sup>26</sup> (provided by J. E. Walker). The expressed protein was purified using a Ni-NTA column (Qiagen) under native conditions. The protein was further purified by gel electrophoresis, and 1.2 mg of purified protein was used to make rabbit polyclonal antisera (Harlan Sera-Lab).

**Western blotting.** Western blots of total protein extracts from *E. cuniculi* spores, infected and non-infected RK-13 cells were incubated with the rabbit anti-EcmtHsp70 (1:1,000), rat anti-EcNTT1-peptide (1:400), rat anti-EcNTT2-peptide (1:3,000), rat anti-EcNTT3-peptide (1:1,000) or rat anti-EcNTT4-peptide (1:2,000) antisera, followed by secondary anti-rabbit or anti-rat antibodies conjugated to peroxidase (Sigma). The blots were developed using enhanced chemiluminescence (Amersham Biosciences).

**Immunolocalization of proteins in *E. cuniculi*.** *E. cuniculi*-infected RK-13 cells were grown on coverslips until confluent and fixed in 50:50 acetone:methanol (v/v) at  $-20^{\circ}\text{C}$  for 2 h. Cells of *E. cuniculi* were also isolated<sup>17</sup> from lysed infected RK-13 cells. After discarding the liquid growth medium, the material from three 175 cm<sup>2</sup> flasks of highly infected (>70%) RK-13 cells was trypsinised, resuspended in medium, transferred to a 50 ml tube and washed twice with  $1\times$  PBS. The cells were centrifuged at 3,000g for 5 min and resuspended in 15 ml of 0.1% saponin and 0.05% (v/v) Triton X-100 in  $1\times$  PBS. The cells were sheared using a Dounce glass homogenizer (clearance 0.062–0.0875 mm) and passed through glass wool plugs (3–4 ml, 2 g per plug) in 10 ml plastic syringe barrels. The filtered fractions were examined by light microscopy, and fractions contaminated with host cells were discarded. The remaining fractions, which contained a mixture of sporonts, sporoblasts and spores, were fixed as above and attached onto poly-L-lysine-coated slides. After blocking with 5% milk-PBS, the cells were incubated at room temperature ( $\sim 22^{\circ}\text{C}$ ) in 1% milk-PBS and a 1:200 dilution of the rabbit antisera against EcmtHsp70, followed by incubation with goat anti-rabbit secondary antibody conjugated to Alexa-fluor 594 (Molecular Probes). For co-localization experiments, the cells were further incubated with rat anti-EcNTT1-peptide (1:100), rat anti-EcNTT2-peptide (1:200), rat anti-EcNTT3-peptide (1:200) or rat anti-EcNTT4-peptide (1:200) antisera, followed

by a goat anti-rat secondary antibody conjugated to Alexa-fluor 488 (Molecular Probes). Coverslips were mounted with DAPI-containing anti-fade mounting reagent (Vectashield), and observed under a laser scanning confocal microscope (Leica, TCS SP2 UV) fitted with a  $\times 63$  objective (Plan Apo, 1.32 nA). Images were collected using LCS V2.61 software (Leica Microsystems, GmbH) software and processed with Adobe Photoshop CS2.

**Fixation of cells for electron microscopy.** Material from three 75 cm<sup>2</sup> tissue-culture flasks of highly *E. cuniculi*-infected rabbit kidney cells were collected and fixed using one of three fixation protocols: 0.5% glutaraldehyde in PBS for 30 min; 4% paraformaldehyde/0.1% glutaraldehyde in PBS for 30 min; or 4% paraformaldehyde in PBS for 30 min. The cells were then washed and stored in PBS at  $4^{\circ}\text{C}$  and processed for electron microscopy imaging as described<sup>3</sup>.

***E. cuniculi*-transporter-mediated ATP uptake in *E. coli*.** Each *E. cuniculi* transporter was cloned into the expression vector pET16b (Novagen). The constructs were confirmed by sequencing and transformed into *E. coli* DH5 $\alpha$  cells (Invitrogen). For uptake assays, *E. coli* Rosetta 2(DE3) pLysS cells (Novagen) were transformed with purified recombinant plasmids and grown to an  $A_{600}$  of 0.6. Expression of the transporter was induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After 1 h, the cells were sedimented (3,000g, 5 min,  $4^{\circ}\text{C}$ ), resuspended in 50 mM potassium phosphate buffer (pH 7.5) to an  $A_{600}$  of 5.0, and used for all uptake experiments.

To analyse the specificity<sup>6,11</sup> of the *E. cuniculi* transporters, 100  $\mu\text{l}$  of induced *E. coli* cells containing either a recombinant plasmid encoding EcNTT1, EcNTT2, EcNTT3 or EcNTT4, or pET16b alone as a control, were added to 100  $\mu\text{l}$  of incubation medium (50 mM  $\text{KPi}$  (potassium phosphate, inorganic,  $\text{KH}_2\text{PO}_4$  and  $\text{K}_2\text{HPO}_4$  mixed to given pH) buffer, pH 7.5) containing 2 nM [ $\alpha$ -<sup>32</sup>P]-labelled ATP with 1  $\mu\text{M}$  cold ATP (Perkin Elmer). Uptake of the [ $\alpha$ -<sup>32</sup>P]-ATP was at  $30^{\circ}\text{C}$  with constant stirring, and was terminated by the addition of 2 ml ice-cold  $\text{KPi}$ , followed by rapid filtration through cellulose nitrate filters (0.45  $\mu\text{m}$  pore size). The filters were washed once with 2 ml ice-cold  $\text{KPi}$  and were transferred to a vial for scintillation counting. All experiments were performed as independent triplicates starting from individual colonies of bacteria. Estimation of the kinetics of ATP transport and back-exchange transport assays were performed as described previously<sup>6,11</sup>.

To determine the kinetic and parameters of ATP transport by EcNTT1–EcNTT4, the external substrate concentration was varied in the range from 0  $\mu\text{M}$  to 128  $\mu\text{M}$  and initial transport rates were determined (0–15 min). The data were fitted with the Michaelis–Menten equation to determine the  $V_{\text{max}}$  and apparent  $K_{\text{m}}$  of transport by iteration. The competition experiments were carried out with 50,000-fold excess of external substrate. For back-exchange experiments, *E. coli* cells expressing the recombinant EcNTT3 were incubated for 20 min in  $\text{KPi}$  containing 1  $\mu\text{M}$  radio-labelled [ $\alpha$ -<sup>32</sup>P]-ATP. The cells were collected by centrifugation at 3,000g for 5 min, and were washed and resuspended in  $\text{KPi}$  with, or without, non-labelled ATP or ADP at a final concentration of 250  $\mu\text{M}$ . Back-exchange experiments were carried out at  $30^{\circ}\text{C}$  and were terminated at defined time intervals by rapid filtration over cellulose nitrate filters. The filters were washed once with 2 ml ice-cold  $\text{KPi}$  and transferred to a vial for scintillation counting. Experiments were performed as independent triplicates.

26. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).

# naturejobs

**JOBS OF  
THE WEEK**

**T**he past few months have seen a host of headlines hinting that interest in energy research is likely to continue for years to come. Each story has implications for expanded career opportunities. In March, Tesla Motors, a company in San Carlos, California, founded and financed by Internet entrepreneurs, began production of its high-performance electric car. The move suggests that the US car industry, which has often been slow to innovate, is at last opening up to ideas from other technological disciplines. Hot on the heels of the Tesla stories, an article in *The Wall Street Journal* mentioned that large automotive companies are beginning to seek licences for technologies — in much the same way that big pharmaceutical companies buy intellectual property from biotechnology firms and universities.

And this month, the US Department of Energy released a report predicting that the country could produce 20% of its power via wind by 2030. Denmark, meanwhile, the world leader in wind production of energy by percentage, has been reported as seeking to generate 50% of its power by wind in 2025, up from 20% currently.

Then there are the three viable US presidential candidates, who have all put forward policies to reduce carbon emissions. This attention to climate change marks not only a reversal from President Bush's stance, but a route towards investment in a suite of renewable energies. Meanwhile, the numerous political, economic and societal pressures are driving energy research. These include record-high petrol prices, sky-rocketing levels of car ownership in India and China, fears of the impact of global warming, and a need to reduce reliance on oil produced in politically sensitive areas such as the Middle East and Russia.

The headlines also suggest that private investment is increasing. If public investment follows, we can expect to see more scientists in more disciplines — from materials science to biosciences — dip their hands into what should be a growing alternative-energy funding pot. Political and environmental motivations mean a bright future for energy research and energy-focused careers.

**Paul Smaglik moderates the *Naturejobs* Nature Network career-advice forum.**

## CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**

Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)  
**Southwest UK/RoW:** Nils Moeller (4953)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)  
**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)  
**North Germany/The Netherlands:**  
Reya Silao (4970)  
**South Germany/Austria:**  
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:** Dennis Chu

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718

Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**  
Vikas Chawla (+91 1242881057)  
e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe (+81 3 3267 8765)  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)  
**Business Development Manager, Greater China/Singapore:**  
Gloria To (+852 2811 7191)  
e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# Westernizing Eastern-bloc science

Scientists are learning how to use what the EU offers, says **Quirin Schiermeier**.

**W**hen the Czech Republic joined the European Union (EU) in 2004, its research and higher education were still in transition from its old Eastern-bloc system to a Western structure based on grants, competition and peer review. Its scientific output lagged far behind the EU average. But the picture is very different now.

"Things have improved greatly here," says Vaclav Horejsi, director of the Czech Academy of Science's Institute of Molecular Genetics in Prague. "Fifteen years ago we were really scared of losing all the best young people to the West. Now many expatriates are returning with loads of invaluable contacts and experiences that often open the doors to European networks and research collaborations."

Not only is domestic funding getting better — with the Czech Republic putting more into science than some long-time EU members (see graph) — but some teams and institutes are becoming sought-after partners in European research collaborations.

Its former Eastern-bloc neighbours, however, have had varied fortunes since joining the western European community. Slovakia, Slovenia, Estonia, Hungary, Latvia, Lithuania and Poland joined the EU in 2004, followed by Romania and Bulgaria last year.

University education and research were traditionally separate in Eastern-bloc countries, with most research carried out in institutes run by national academies of sciences. There was little or no competition for grants because academy institutes received a lump sum from the government. In most of the new EU states the academies still exist, with ageing staff and strong resistance to competition through grant-based funding.

None of the new members is yet up to standard in the number of grant proposals submitted, in industry participation (all EU consortia include partners from industry) and in the number of principal investigators in charge of EU research consortia. But some scientists are getting to grips with the new funding opportunities such as the EU's Seventh Framework Programme (FP7), running from 2007 to 2013. About 10% of successful proposals come from the new member states, according to preliminary results of the first calls for proposals.

Initial FP7 figures show that they are as successful as anyone else in the evaluation process. In fact, the overall success rate of Czech teams in the first 57 calls was above the EU average in most areas. Those from Estonia and Lithuania did similarly well. They and, according to early indications, Hungary are doing well in a number of fields, such as biotechnology.

A small but growing number of competitive teams and institutes is leading the way. In biotechnology, nanosciences, materials research and aeronautics, Czech teams have obtained about 1.5% of the €17.5-billion (US\$27-billion) framework programme.

Prague's Institute of Molecular Genetics is located in one of central Europe's most modern biotech campuses. Six junior group leaders, all Czech- and



Slovak-born scientists recruited from abroad, have recently taken up work at the new €35-million building with its animal house, conference hall and state-of-the-art lab equipment.

"The chance to build my own group was the main motivation for me to return," says cell biologist David Staněk, who completed postdoctoral research at the University of Washington in Seattle and at the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) in Dresden, Germany, before returning home in 2005. "In a small country like mine, individual scientists have more influence on the course of things than they would have in the United States or Germany."

The Prague institute is involved in numerous joint projects with partners in western Europe, including five EU-funded collaborations. Horejsi's group is preparing an FP7 grant proposal for a project in molecular immunology, coordinated by the MPI-CBG.

This large institute, only a short train-ride from Prague, has become an important training ground for young scientists from central and eastern Europe. The European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, is equally valuable. Both are focal points for successful collaborations — one of the keys to success for the new states.

"For a Max Planck researcher it is the most natural thing in the world to ring up a colleague at Imperial College London. It's just built in their psyche for years," says Brendan Hawdon, a European Commission official who oversees the coordination of FP7.

**"The basis of our success is the young, mobile generation of scientists with networks abroad"**  
— Mart Ustav



The new member-states' institutions have less of a tradition of scientific cooperation. Hawdon admits that there is snobbery, too. Researchers in rich countries are sometimes reluctant to consider lesser-known scientists from weaker countries as EU-funded partners, he says.

But as EU-funded research is primarily aimed at scientific excellence, and not at regional convergence, the commission will choose the best proposals regardless of whether they come from teams in Oxford, Prague or Sofia. "Things must happen naturally," says Hawdon. "There will be no forced marriages."

### Radically modernized system

Estonia, with only about 3,000 researchers among its population of 1.3 million, is well on the way to becoming a model for the Baltic region. "We do know our weaknesses," says Mart Ustav, department chair of biomedical technology at the University of Tartu. "Public expenditure on science is still rather low, and industry doesn't add much to it."

But since Estonia gained independence from the Soviet Union in 1991, it has changed its science system more radically, and more successfully, than any other formerly communist country in Europe. The merger in 1993 of the Academy of Sciences with the country's universities in Tallinn and Tartu ended the separation between research and teaching, and generated a situation in which students and young scientists can do active research in the lab early on.

In addition, the government invested to the best of its ability in new institutes and research equipment. This policy has paid off. The Institute of Technology at the University of Tartu has become a scientific powerhouse, involved in many transnational collaborations and with a strong scientific output. The country's overall return rate from FP7 is at least three times higher than its financial contribution to the programme. "We would be happy if our example stands up," says Ustav.

Part of the Estonian strategy is to offer expatriates who are willing to return the same excellent conditions they have got used to abroad — an idea that the tiny nation can more easily put into action



**"There's no lack of talented students, but they're not aggressive enough in getting out and letting us know what they're good at" — Ivan Dikic**

than its larger neighbours such as Poland with their more cumbersome political systems.

"The basis of our success is the young, mobile generation of scientists with networks abroad," says Ustav. But once you're used to "driving a Mercedes," he says, you just don't want to steer an old clunker any more.

### The challenge of competition

Although a few top institutes, from Budapest to Tallinn, do now attract a number of EU grants, the large majority of researchers in the new member-states show little enthusiasm for European research, instead looking for support elsewhere. Many don't even try to get European money. They may see no need, no hope of succeeding or no way to grasp the bureaucratic system.

"Let's face it, the 'old guard' is simply not competitive enough," says Croatian-born Ivan Dikic, a biochemist at the University of Frankfurt in Germany, referring to middle-aged tenured professors. Because few labs are doing top research, and those labs don't cover all fields, it remains difficult to find scientific partners in the new member states. "There is no lack of talented students and excellent young scientists in the region, but often we just don't know who they are," he says. "They are not aggressive enough in getting out, promoting their skills, and letting us know what they're good at."

Poland, by far the largest of the new member states, isn't performing well. Its teams are underrepresented: although it has 8% of the total EU population, it takes part in only 2.3% of successful EU-funded cooperation programme proposals.

Marta Miaczynska, now a group leader at the International Institute of Molecular and Cell Biology (IIMCB) in Warsaw, completed postdoctoral research at both the MPI-CBG and EMBL before returning to her native Poland three years ago. The contacts and personal reputation she built up while working abroad helped her establish her new group as one of the 12 partners participating in an €11-million EU-funded collaboration on cellular signalling and endocytosis.

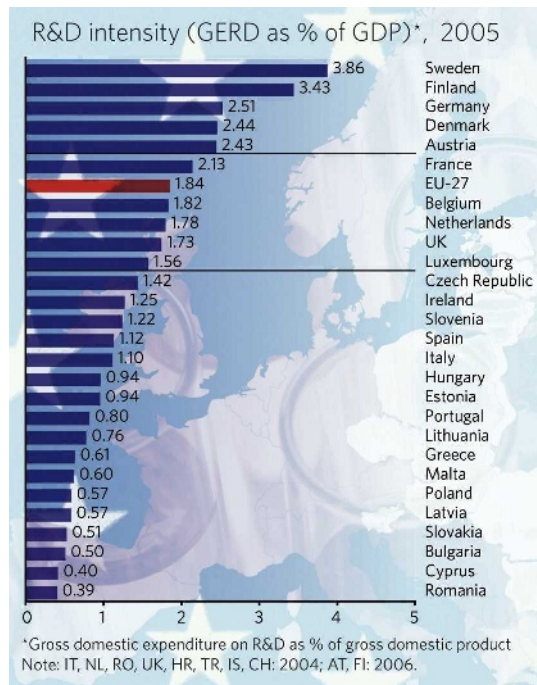
But the IIMCB, governed by an international advisory board, is an exception. The low success rates of Polish teams in the first FP7 calls — and Poland's total failure in the first round of applications for young investigator awards by the European Research Council — led the government to propose a sweeping reform of the system, with calls for more competition, more grant money, a better retirement scheme for ageing professors in permanent positions, and the abolition of *Habilitation*, the second thesis required to become a professor.

"We need brain circulation instead of brain drain," says Barbara Kudrycka, the Polish minister of science and higher education (see page 438). Fresh money and more domestic competition for grants would help, says Miaczynska. But scientists must seek EU funding more aggressively as well. "People here are still learning that you need to fight very proactively for grants, particularly if you're lesser known and lack good contacts in the West," she says. "The EU bureaucracy can be scary. But often people are just not motivated enough, thinking it's too much work and too competitive anyway."

Experts think it may take 10 years or so before the new members will be on an equal footing, scientifically, with most western European countries — assuming sound investment and forward-looking policies.

**Quirin Schiermeier is Nature's German correspondent.**

SOURCE: DG RESEARCH; DATA: EUROSTAT/OECD



# MOVERS

**Frank Torti, chief scientist, US Food and Drug Administration**



**1993–present:** Director, Comprehensive Cancer Center, and Chair, Department of Cancer Biology, Wake Forest University School of Medicine, Winston-Salem, North Carolina

**1986–93:** Associate professor, Stanford University Medical Center, Stanford, and Chief, Oncology Section, VA Medical Center, Palo Alto, California

Frank Torti is excited about taking on what promises to be an onerous job. As the first chief scientist at the US Food and Drug Administration (FDA), he is eager to help the embattled organization use the latest research and technology to create more rigorous and efficient regulatory controls for approving new drugs. "I think I can help them break new scientific ground," he says.

After earning a biology degree from Johns Hopkins University in Baltimore, Maryland, Torti pursued an MD at Harvard Medical School. An interest in the molecular biochemistry of nutrition then led him to do a Master's in public health at Harvard. But, motivated in part by his parents' battle against cancer, he eventually accepted a fellowship in oncology at Stanford University.

Torti went on to create one of the first clinics treating genital and urinary cancer to bring together radiation therapists, medical oncologists and surgeons. And he developed chemotherapy regimens for bladder and prostate cancers that became standards of care. As executive director of the Northern California Oncology Group, Torti also learned the inner workings of trial design and patient recruitment.

But he missed the laboratory and so made an unusual move, temporarily resigning his faculty position to become a visiting scholar in Stanford's pharmacology department. There, he discovered basic molecular mechanisms underlying the regulation of proteins involved in iron metabolism, and how these are modified by cancer.

Torti jumped at the chance to lead both Wake Forest's Comprehensive Cancer Center and its basic science department in cancer biology. There, his success in building clinical trials and training programmes brought widespread recognition, and his basic-science research led to a prestigious National Institutes of Health MERIT award. "Frank brought Wake Forest's Cancer Center from the backwater to be a major player," says colleague Steve Akman. He predicts that Torti will help the FDA to recapture lost esteem by recruiting top talent and organizing the agency's responsibilities among its constituencies.

Torti says he wants to act as an advocate for the science community. He also wants to integrate cross-cutting themes — such as nanoscience and toxicogenomics — throughout the agency. And he plans to develop a top-notch fellowship training programme, hoping to make the agency more attractive to bright young scientists interested in translating basic science into clinical practice.

**Virginia Gewin**

## NETWORKS & SUPPORT

### Upping student numbers and diversity

Attracting undergraduates to science is an ongoing challenge, particularly for small liberal-arts institutions that lack access to federal research dollars. But once they engage students, smaller schools focused on teaching create a surprisingly strong source of scientists for graduate schools. In April, the Howard Hughes Medical Institute (HHMI), a non-profit philanthropic biomedical organization, announced a grant of \$60 million to be split between 48 such undergraduate institutions to create innovative ways to engage students in the biological sciences.

Although this funding programme has been in place since 1988, Peter Bruns, HHMI vice-president for grants and special programmes, says that the focus this year has been on diversity. Bruns says that the HHMI specifically sought to capture a mix of ethnic, gender and academic backgrounds in this year's awardees. More than one-quarter were first-time HHMI grantees. And to bolster the number of historically black colleges receiving HHMI monies, which has declined in recent years, the institute held a pre-competition workshop to review proposal particulars. Five historically black colleges were funded.

One of those was North Carolina Central University (NCCU) in Durham. At present, only 5% of NCCU's

students major in science. With HHMI funds, NCCU will craft summer lab-based research programmes for middle and high schools, and link them to its existing undergraduate research and mentoring opportunities. "We will pave a 10-plus-year path from middle school to college and graduate professional schools," says Gail Hollowell, the university's HHMI programme co-director.

In a bid to reverse a recent 20% to 15% dip in science majors, Drew University in Madison, New Jersey, is using its HHMI funds to capitalize on regional assets that introduce students to real-world scientific challenges. It has devised 'science and society' seminar courses and symposia featuring executives from Wyeth Pharmaceuticals.

Bruns says that the HHMI funding has an added bonus: awardees comprise a *de facto* network. Meetings of HHMI programme directors from different schools are often the genesis for additional grants. For example, Davidson College in North Carolina received additional HHMI funds to put in place a full-service microarray data-analysis infrastructure — allowing fellow undergraduate colleges to conduct high-tech experiments at lower cost.

**Virginia Gewin**

#### POSTDOC JOURNAL

### An impassive observation

This morning I watched a monkey named Bubba viciously attack fellow unit member Meena. She screamed in fear and fled to the bulk of her erstwhile lover, who just cowered, pretending to shield her. Meena did sink her teeth into Bubba, but she came off second best, with blood dripping from her arm. Nobody in the unit had come to her aid. I noted the events on my palmtop, a seemingly cool, detached observer. And I wished Bubba a slow, painful death by leopard mauling.

We are trained never to anthropomorphize when interpreting animals' behaviour. We are trained to be unbiased and unemotional in our reporting. I agree with this. But I wonder, are we hiding one of our human strengths? I throw myself into my work physically, mentally and emotionally. I think it's the emotional investment that makes me a meticulous scientist; after all, it is my fondness for the animals that leads me to search for hours to find them. And I find myself driven to scrutinize the subtle and overt actions of my study subjects.

Many biologists know their subjects as individuals, not just numbers on a data sheet. This enhances their ability to understand and interpret those subjects' interactions. As a young scientist I get the impression that we have to hide this. I don't want to plead for 'emotional' reporting in peer-reviewed journals. But I do want to acknowledge that we can be both emotionally involved and objective. And this is a good thing.

**Aliza le Roux is a postdoctoral fellow in animal behaviour at the University of Michigan.**





Image Source: Wikimedia Commons

**Naturejobs is pleased to present the**

# SPOTLIGHT ON ALABAMA

**ISSUE DATE:**

**JUNE 5, 2008**

**DEADLINE FOR ADVERTISERS:**

**MAY 29, 2008**

The upcoming *Naturejobs* Spotlight will raise awareness of Alabama's rapidly emerging biotech industry. This special feature in *Nature* offers Alabama research institutions and organizations a unique opportunity to recruit top-tier scientific talent while building their reputation among the local and international scientific research community.

Alabama is home to seven research universities and over 90 life sciences companies that raised \$104 million in capital and generated \$622 million in total sales in 2006.\* The state has produced six FDA approved cancer treatments with six additional drugs currently in clinical trials at Southern Research, based in Birmingham, Alabama.\*

Over the past two years, there have been new additions to the Alabama life sciences landscape including the Shelby Interdisciplinary Biomedical Research Building, the HudsonAlpha Institute for Biotechnology, and the Auburn University Research Park. These impressive new facilities are evidence of Alabama's rapidly growing biotech industry.

The Spotlight on Alabama will be a valuable reference for researchers, students, and educators searching for career opportunities and events in Alabama and the surrounding area. It will be eagerly read by those drawn to the front pages of *Nature's* recruitment section.

Job advertisements will also receive a complimentary 8-week posting on naturejobs.com and will be matched to all relevant content across nature.com. Event advertisements will receive a complimentary 8-week posting on natureevents.com – the events portal for *Nature*.

**TO PROFILE YOUR ORGANIZATION AND EVENTS TO POTENTIAL EMPLOYEES, PARTNERS AND INVESTORS, PLEASE CONTACT:**

**Helen Wu | T: 212-726-9347 | F: 800-989-7103 | E: [h.wu@natureny.com](mailto:h.wu@natureny.com) | Naturejobs, 75 Varick Street, 9th Floor, New York, NY, 10013, USA**

\*Source: The official website of the State of Alabama (<http://www.alabama.gov>), 2007.

IN12857R

**naturejobs**

nature publishing group **npg**





## NATUREJOBS PRESENTS A HIGHLIGHT ON SPAIN

**ISSUE DATE: 10TH JULY 2008**

**ADVERTISING DEADLINE:**

4th July 2008

16.00 CET / 15.00 GMT

*Nature* has unmatched impact and reach; this Highlight, exclusively dedicated to promoting the strong scientific environment in Spain, receives significant exposure.

This particular issue of *Nature* will have a bonus distribution at ESOF in Barcelona and will also feature a career and recruitment editorial piece on 'Integrity in Science'.

Investment in Spanish science is on the rise. Though the country has always produced quality scientific research, increased funding from national and local governments, and new research centres have created a scientific landscape which attracts an increasing number of outstanding researchers.

The *Naturejobs*' Highlight provides Spanish companies and scientific institutions with the opportunity to profile themselves as employer-of-choice, and attract world-class researchers and scientists for their projects. The Highlight also serves to showcase centres of excellence, scientific hubs in Spain, and international research projects to *Nature*'s global audience in order to generate awareness and position themselves in the international scientific arena.

Whether you're recruiting, generating awareness, or positioning your research group, this feature provides the competitive edge required to thrive in the global scientific community and is the best opportunity to deliver targeted impact.

**To find out how to take part in this opportunity, please contact:**

**Evelina Rubio- Hakansson**

*Naturejobs*

London Office

T. +44 (0)20 7843 4973

E. e.rubiohakansson@nature.com

**Claudia Paulsen Young**

*Natureevents*

London Office

T. +44 (0)20 7014 4015

E. c.paulsenyoung@nature.com

**NATUREJOBS@NATURE.COM**

**WWW.NATUREJOBS.COM**

**naturejobs**

nature publishing group **npg**

# The Neanderthal correlation

A question of breeding.

**Jeff Hecht**

"You don't really want to know," Beth said, her eyes glancing up from her messy desk to the clock and back down without meeting mine.

"I do want to know," I insisted. After three years of correlating the reconstructed Neanderthal genome with modern human populations, she had to have found something interesting. The idea had sounded great when she had suggested it, and the grant committee had loved my proposal. But with the final report ten months overdue, they wouldn't approve any new proposals.

"You'll wish you hadn't said that if I tell you," she said, staring down. "People are going to be upset about this."

"I want to know," I insisted. What I really wanted was a report to drop on the contract officer's desk, but saying that hadn't worked the last time. Or the time before that. Beth gets too deep into her research. I run a big human-genetics lab. I deliver results; I don't invest my ego in big-picture hypotheses or in worrying why the Neanderthals died out. "I don't have an oar in the debate over whether or not the Neanderthals interbred with Cro-Magnons. I just want to know what the data say."

"Really?" she asked, not looking convinced at all. "You said that the last time, remember?"

I didn't. I tried to ignore her obsessions. "Please," I asked.

"It's not just you. Nobody is going to want to hear this. Believe me, John, believe me."

"I'm a scientist. I want to know the truth!" More importantly, I wanted to finish the contract; that was my job as principal investigator. I'd always succeeded before; that was why after two decades at the university I was department chair and Beth was still a research assistant.

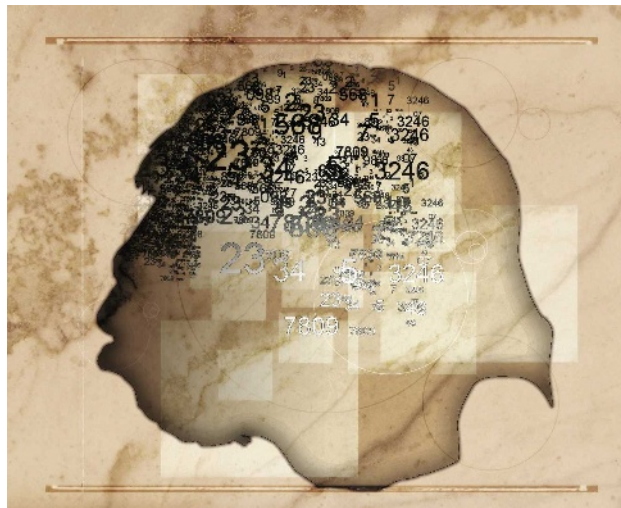
"Are you sure?" Beth asked, looking a little less uncertain.

"Yes," I said, trying to smile. "I know you've got something very interesting to tell me." That sometimes worked.

She nodded, her usually expressionless face showing a shadow of a smile. "I found strong genetic correlations between Neanderthals and modern subpopulations," she said. "A lot more than I had expected."

"What about correlations coming from the last common ancestor?" That was the safe correlation. Sapiens and Neanderthals had split around 800,000 years ago, so they had to share lots of genes that chimps didn't have.

"Some are," she said. "They're easy to find because they're in all modern populations. These genes are present in only some modern subpopulations, and the



statistics show only about 25,000 years of divergence between the Neanderthals and Sapiens. That has to be interbreeding. The earlier studies had missed it because they hadn't considered the changing impact of natural selection over time."

"You can back that up?"

"Absolutely," Beth was always meticulous about her data.

I didn't have to force a smile. "That's fascinating," I said. "It will make *Nature* for sure." It would get a lot of people hot under their collective collars, but that was fine. Evidence of interbreeding with Neanderthals would create a new paradigm for hybridization being behind the rapid advance of modern humans and make me famous. "What genes are involved?"

"That's the surprise," Beth said, and she smiled so broadly that she looked almost attractive despite her unkempt red-grey hair and nondescript clothes.

"Oh?"

"The genes for red hair and pale skin didn't match well enough to show a correlation, but I found a correlation for genes linked to other traits. There's a gene cluster linked to advanced mathematics skills, information processing, logic, analytical intelligence, concentration skills, obsession-compulsion and Asperger's

syndrome. That cluster correlates very strongly. I can trace some genes back to the interglacial around 450,000 years ago, and others back to another burst of evolutionary innovation during the Eemian interglacial about 130,000 years ago." She rambled on with endless details.

Something wasn't right. She was linking genes for advanced mental skills to Neanderthals. "I'm confused," I said when she paused for a breath. "You're correlating genes linked to modern human intelligence with Neanderthal populations. What am I missing?"

"You didn't want to hear me, I knew that."

"No, I want to hear you. I just asked a question."

"You don't, because I already told you."

I looked at Beth blankly, realizing I was missing a key part of the puzzle. "You said these were Neanderthal genes?"

"Yes, they were," she said. "They weren't in the modern human genome until Neanderthals interbred with Cro-

Magnons between 25,000 and 30,000 years ago."

"Advanced mathematical processing? Shouldn't that have been missing from the Neanderthal genome?"

"No, I found that Neanderthals lacked genes linked to successful socialization and management skills. They could count perfectly well, but they couldn't deal with groups. Socialization genes came from Sapiens"

"You're trying to tell me..." I said, but my mental censor blocked the idea.

"That human mathematical intelligence came from Neanderthals? That's what the data say. The Cro-Magnons had the social skills. But that isn't all."

I stared at her. I couldn't tell that to the research council.

As usual, she couldn't read the warning look on my face. "The hybridization was successful in the Stone Age, but the environment has changed. I found that modern culture selects for socialization but against the Neanderthal traits for mathematics and intelligence," she said, and looked down. "I don't know how you'll survive when our genes are gone."

**Jeff Hecht ([www.jeffhecht.com](http://www.jeffhecht.com)) is Boston correspondent for *New Scientist* and contributing editor to *Laser Focus World*.**

JACEY